R E P O R T   R E S U M E S

ED 020 499 AL 001 073

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION--II. PROGRESS REPORT 2, SEPTEMBER 1, 1961 TO FEBRUARY 1, 1962.

BY- LANE, HARLAN AND OTHERS
MICHIGAN UNIV., ANN ARBOR,COLL.OF LIT.,SCI.,ARTS
REPORT NUMBER ORAP-04411-2-P PUB DATE MAR 62
CONTRACT OEC-SAE-9265
EDRS PRICE MF-~~$0.75~~ $1.00 HC-$7.92 196P.

DESCRIPTORS- *ORAL COMMUNICATION, *PERCEPTION, AUDITORY PERCEPTION, *LANGUAGE RESEARCH, *SPEECH, AURAL STIMULI, VOWELS, REINFORCEMENT, PERSONALITY STUDIES, ACOUSTICS, PSYCHOLINGUISTICS, EXPERIMENTAL PROGRAMS,

THIS DOCUMENT REPORTS THE PROGRESS FROM SEPTEMBER 1, 1961 TO FEBRUARY 1, 1962 OF AN EXPERIMENTAL PROGRAM IN SPEECH CONTROL. THE TOPICS DISCUSSED ARE--(1) PARAMETERS OF VOWEL PERCEPTION, (2) RELATIONS AMONG RESPONSE RATE, TOPOGRAPHY, AND SCHEDULES OF REINFORCEMENT--METHODS AND FINDINGS IN AN ANALYSIS OF THE VOCAL OPERANT, (3) FOREIGN ACCENT AND SPEECH DISTORTION, (4) METHODS FOR SELF-SHAPING ECHOIC BEHAVIOR, (5) EFFECTS OF TRAINING ON ESTIMATES OF VOWEL LOUDNESS, (6) JUDGMENT OF PERSONAL CHARACTERISTICS AND EMOTIONS FROM THE NONVERBAL PROPERTIES OF SPEECH, AND (7) PERSONALITY STEREOTYPES IN VOICE--A RECONSIDERATION OF THE DATA. (SEE RELATED DOCUMENTS ED 003 883 THROUGH ED 003 887 AND AL 001 072.) (DO)

(No Bureau number)

04411-2-P

PA 48

THE UNIVERSITY OF MICHIGAN

ED020499

COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS

DEPARTMENT OF PSYCHOLOGY

Progress Report 2

September 1, 1961, to February 1, 1962

Experimental Analysis of the Control of Speech Production and Perception: II

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Under Contract with:

Language Development Section
U. S. Office of Education
Contract No. SAE-9265

Administered through:

March 1962

OFFICE OF RESEARCH ADMINISTRATION · ANN ARBOR

[illegible]L 001 073

THE UNIVERSITY OF MICHIGAN
COLLEGE OF LITERATURE, SCIENCE, AND THE ARTS
Department of Psychology

Progress Report 2
September 1, 1961, to February 1, 1962

EXPERIMENTAL ANALYSIS OF THE CONTROL OF SPEECH PRODUCTION AND PERCEPTION: II

ORA Project 04411

under contract with:

LANGUAGE DEVELOPMENT SECTION
U.S. OFFICE OF EDUCATION
CONTRACT NO. SAE-9265

administered through:

OFFICE OF RESEARCH ADMINISTRATION ANN ARBOR

March 1962

CONTENTS

RESEARCH COMPLETED

PARAMETERS OF VOWEL PERCEPTION

THE RELATIONS AMONG RESPONSE RATE, TOPOGRAPHY, AND SCHEDULES OF REINFORCEMENT: METHODS AND FINDINGS IN AN ANALYSIS OF THE VOCAL OPERANT

FOREIGN ACCENT AND SPEECH DISTORTION

METHODS FOR SELF-SHAPING ECHOIC BEHAVIOR

EFFECTS OF TRAINING ON ESTIMATES OF VOWEL LOUDNESS

THE JUDGMENT OF PERSONAL CHARACTERISTICS AND EMOTIONS FROM THE NONVERBAL PROPERTIES OF SPEECH

PERSONALITY STEREOTYPES IN VOICE: A RECONSIDERATION OF THE DATA

RESEARCH IN PROGRESS

REINFORCEMENT OF COVERT VOCAL RESPONDING

ON THE PERCEPTION OF AFFECT IN SPEECH

CONTROL OF AUDITORY ATTENTION

SOME PROPERTIES OF COVERT CHAINING IN LEARNING TO IDENTIFY STIMULI

THE PERCEPTION OF AFFECT IN A FOREIGN LANGUAGE AS A FUNCTION OF SECOND-LANGUAGE TRAINING

APPENDIX

STAFF

SPECIAL ACTIVITIES

DISTRIBUTION

RESEARCH COMPLETED

PARAMETERS OF VOWEL PERCEPTION

THE RELATIONS AMONG RESPONSE RATE, TOPOGRAPHY, AND SCHEDULES OF REINFORCEMENT: METHODS AND FINDINGS IN AN ANALYSIS OF THE VOCAL OPERANT

FOREIGN ACCENT AND SPEECH DISTORTION

METHODS FOR SELF-SHAPING ECHOIC BEHAVIOR

EFFECTS OF TRAINING ON ESTIMATES OF VOWEL LOUDNESS

THE JUDGMENT OF PERSONAL CHARACTERISTICS AND EMOTIONS FROM THE NONVERBAL PROPERTIES OF SPEECH

PERSONALITY STEREOTYPES IN VOICE: A RECONSIDERATION OF THE DATA

PARAMETERS OF VOWEL PERCEPTION

Harlan Lane
The University of Michigan

Parameters of Vowel Perception

Harlan Lane
Communication Sciences Laboratory
University of Michigan[1]

There are two simplifying assumptions about the perception of complex stimuli, such as speech signals, that must be abandoned. The first is that, with the exception of a few notable illusions, perception is veridical. The second is that the perception of a single attribute of a complex stimulus is invariant under changes in the other attributes.

An example of the first assumption, pertinent to speech perception, is that the loudness of a speech signal is linearly related to its intensity, or that its pitch is linearly related to its fundamental frequency. Such statements imply that the sensory transducers involved have a linear operating characteristic--a simple but not very likely psychophysical law. In fact, the empirical evidence favors a power law description of sensory process along intensive continua (Stevens, 1958); in this event, a linear relation between apparent and physical magnitudes is a very special case. In the area of speech perception, some investigators have taken note of the findings of pure-tone psychophysics, such as the sone scale for loudness (Stevens, 1955), and the mel scale for pitch (Stevens, et al., 1937). It has been suggested, for example, that the mel scale might be a more suitable coordinate for the sound spectrogram than the frequency scale currently in use (Fant, 1960). There has been considerable reluctance, however, to extrapolate from "the psychophysical principles of the perceptions of abstract sounds" (Lehiste and Peterson, 1959) to the case of speech signals, and these principles generally have not been applied.

Allowing that the listener's perception of speech may not be linearly related to speech parameters, it should be recognized that the speaker's perception of his own speech, his autophonic output, may be similarly complex. Furthermore, reception scales and autophonic scales may be non-linearly related not only to physical magnitude but also to each other (Lane, et al., 1961). Without a knowledge of these scales, the total verbal episode presents some curious anomalies. For example, a subject is asked to faithfully reproduce a disyllable whose segments have an intensity ratio of ten decibels; he responds reliably with a stress difference of three decibels. The puzzle is solved when we realize that the subject can only equate the apparent magnitude of the stimulus and of his response. Since the subjective scales of speech loudness and of autophonic level are not the same, the magnitudes reported equal will not be equal in physical units (Lane, 1961a).

The second assumption cited, that of a one-to-one relation between parameter and percept, has the following forms apropos of speech perception: the loudness of a speech signal is unaffected by its pitch; the pitch of a vowel is unaffected by its formant frequencies; etc. When the terms loudness and intensity, or pitch and fundamental frequency, are used interchangeably, this may be evidence of both assumptions, veridicality and invariance of perception, operating in concert. Students of speech perception have given cursory recognition to the interaction effects of stimulus parameters observed in pure-tone psychophysics, such as the relations among intensity and frequency described by the equal-loudness contours. One author has suggested that the design of a speech intensity meter might include a weighting function, applied to the frequency spectrum of the signal to be measured, in accordance with equal-loudness data (Fant, 1960). Nevertheless, there has been no systematic effort heretofore to chart the relations among intensity, fundamental frequency, duration and their subjective correlates.

Figure 1 shows the guidelines for an analysis of a minimal verbal episode when these simplifying assumptions are qualified. A vowel is spoken and the subject is requested to repeat it accurately (give an echoic response). The present discussion is limited to a consideration of the vowel parameters: sound pressure level (As), duration (Ds), and fundamental frequency (Ps). The intensity of S's response (Ar) is a function, first of all, of the perceived loudness of the stimulus (As_p) which is, in turn, a function of the stimulus parameters As, Ds, Ps. The intensity of S's responses may be expected to vary also as a function of the perceived duration (Ds_p) and perceived pitch (Ps_p) of the stimulus; these are, of course, also functions of the stimulus parameters. Finally, the intensity of S's reponses may depend upon autophonic scales of loudness (Ar_p), pitch (Pr_p), and duration (Dr_p), each of which depends, in turn, on response parameters Ar, Pr, and Dr. A comparable analysis applies to the duration (Dr) and fundamental frequency (Pr) parameters of the matching response. Note that all interaction terms have been omitted from this diagram. For example, the function for As_p may be expanded: $As_p = g(As, Ds, Ps, As \times Ds, As \times Ps, Ds \times Ps, As \times Ds \times Ps)$ and similarly throughout the analysis. Of course, many of the variables that magnify the complexity of this a priori analysis may, on empirical evidence, be neglected without serious error.

The diagram in Fig. 1 serves to enumerate the psychophysical functions that must be determined for the prediction and control of a minimal verbal episode such as that described. The present study represents first efforts to determine the form and parameters of many of these functions (those underscored in the diagram). The validity of these subjective scales is then tested by predicting response parameters in the echoic responding situation described above.

Method

Three categories of psychophysical experiments comprise the research to be reported: (1) Those determining autophonic scales, i.e., scales of the speaker's perception of his own speech parameters, (2) those determining reception scales, i.e., the listener's scales of vowel parameters, and (3) those involving echoic responding, in which subjective scales are validated by the prediction of response parameters.

Autophonic scales

(a) Loudness. The scale of autophonic level, employed in the present study, was reported previously by Lane, et al. (1961) and obtained by the method of magnitude production (see below).

(b) Duration. The method of magnitude production was employed (Stevens, 1958) with 10 subjects, serving individually. The subject was seated in an audiometric room, in front of a microphone, and asked to produce the vowel phoneme/a/with moderate loudness and duration. To the duration of this response the experimenter assigned the numerical value 10. A series of values (2.5, 5, 10, 20, 40) were then named, 12 times each, in irregular order and the speaker was asked to respond to each with a vocal production of proportionate duration. Vocal responding was tape recorded (Ampex 300) and, subsequently, each recorded signal was applied to a calibrated voice-operated relay

(Miratel), whose drop-out time was independent of signal amplitude. The VOR controlled the time interval section of a frequency counter (Hewlett-Packard 522B), which read the duration of the response by counting cycles of a 10 kc frequency standard. A parallel printer (Hewlett-Packard 560A) recorded numerical values.

(c) Pitch. The method of category rating (Stevens, 1958) was employed with nine subjects. The subject was seated in an audiometric room and asked to produce the vowel /a/ with moderate pitch. The transduced signal was sent to a "pitch meter": an array of filters and switching circuitry suitably arranged to extract the fundamental frequency from this complex vowel sound. The final stage of the pitch meter was a frequency meter (Hewlett-Packard 500B) that converted the sinusoidal input to a d-c signal whose voltage was proportionate to the input frequency but independent of amplitude. This voltage was applied to the vertical axis of an oscilloscope, (Tektronics 533) with disabled sweep. Suitable attenuation was introduced so that the reponse with moderate pitch would just center the point of light on the screen. By adjusting the circuit attenuation, and then requiring the subject to center the light, ten fundamental frequencies (120, 130, 140...190, 200, 210) were obtained, five times each, in irregular order. After each response, S rated his pitch on a seven-point scale. The rating device was a narrow rectangular box on which were mounted seven buttons, 1/4 inch diameter, spaced at one-inch intervals. A scale number (one through seven, left to right) was written above each button and above the left and right extreme buttons the words low and high, respectively, were also printed. Category ratings were recorded by a multiple events recorder in parallel with the button-press device.

(4) Pitch-Amplitude relations. This experiment did not determine an autophonic scale; it was designed to measure the relation between pitch and amplitude when the experimenter constrained one parameter and not the other. Each of ten S's sat in an audiometric room with his head taped to a headrest, in front of a microphone, VU meter (Daven) and pitch meter (described above). The experimenter sat behind the subject and controlled an attenuator and electronic switch (Grason-Stadler 829) that sampled S's vocal output. In the pitch measurement phase of the experiment, S watched the effect of his voice on the needle of the VU meter, whose scale had been partially obscured, and his task was to center the pointer on the face of the meter. The experimenter controlled the gain in the microphone circuit and thereby determined the vocal level necessary for centering. Autophonic levels of the vowel /a/ were required at 5 db intervals over a 15 db range; each level was produced three times in irregular order. When S maintained 0 VU ± 1db for one second, E triggered the electronic switch, which permitted a one-second sample of the vocal response to be tape-recorded (Ampex 300). In the amplitude phase of the experiment, (part two for five Ss and part one for the other five), S watched the effect of his voice on a frequency meter whose scale was offset to range from 120 to 220 cps. Five pitch levels, equally spaced over a range of 120 to 220 cps were required three times each in irregular order. When S maintained a steady pitch at the value requested, E triggered the switch and one second of phonation was recorded.

Tape recordings were processed subsequently to determine the pitch and amplitude of each response. Pitch extraction procedures are described above. A numerical record of the fundamental frequency was obtained by sending the selected sinusoid to the "ten period average" section of a frequency counter (Hewlett-Packard 522B). The average duration of the initial ten periods of the vocal response was measured in milliseconds, to two decimals, by the counter and recorded on a parallel print out. The reciprocal of this datum was taken as an estimate of the autophonic pitch. The amplitude of each response was measured by applying the recorded signal to a speech intensity meter. This device introduced linear, full-wave rectification and then filtering with bandwidth 32 cps, integrating time 11 msec.[2] The peak of the d-c output waveform was read by a peak meter (Control Devices PTM7) and recorded by a parallel printer (Hewlett-Packard 560AR).

Reception scales

(a) Loudness. The scale of received speech loudness employed in the present study was reported previously by Lane (1961) and obtained by the method of magnitude estimation (see below).

(b) Duration. In order to obtain a subjective scale for vowel duration, and to determine the effect of the amplitude and fundamental frequency of the signal on estimated duration, these three stimulus parameters were incorporated in a three-dimensional experimental design. The variables and their levels, shown in Fig. 2, were: duration (200, 300, 400, 500 msec.) sound pressure level (50, 60, 70, 80db) and fundamental frequency (100, 120, 140, 220 cps).

To prepare the stimulus tape recording, four autophonic levels of the phoneme /a/ were generated at 10 db intervals by one speaker. A 500 msec sample of each was obtained with an electronic switch (Grason-Stadler 829; rise time 25 msec), controlled by a calibrated interval timer (Grason-Stadler 471). The fundamental frequency of each of these four signals, which was determined (± 3 cps) with a sound spectrograph (Western Electric BTL-2), was 100, 120, 140, and 220 cps. The four gated signals were recorded on magnetic tape and then each one was copied four times, using the electronic switch to obtain sample durations of 200, 300, 400 and 500 msec. The 16 recorded signals were then processed by the intensity meter (supra) and displayed on a recording oscillograph, which was calibrated so that peak speech intensity could be read ±.5 db. Using this record as a guide, various amounts of attenuation were introduced at the output of the playback tape recorder, so that each of the 16 signals was recorded on a second recorder at 0, 10, 20, and 30 db below zero VU.

The 64 stimuli obtained in this manner were presented in irregular order in three successive series to each of ten unpracticed observers. The subject was seated in front of a microphone in an anechoic chamber; he wore a pair of PDR-10 earphones (mounted in MX-41/AR cushions) that had flat frequency response (± 2db) from 50-4,000 cps. The tape recorded stimuli were sent to a transistor earphone amplifier with high signal-to-noise ratio and flat frequency response, and then to S's headset. The playback system was adjusted so that a 1,000 cps tone, recorded at zero VU, would produce a sound pressure level of 80 db in the earphones (measured with a 6 cc coupler, calibrated condenser microphone, and Ballantine rms VTVM). Each series of 64 stimuli began with a 500 msec, 1,00 cps tone recorded at 0 VU that served as the standard. The experimenter assigned the value "100" to the apparent duration of this stimulus, and S was instructed to estimate the duration of each subsequent vowel by assigning it a number in the same proportion to 100 as its apparent duration was to the standard.

(c) Pitch. The category rating procedure used to determine a scale of autophonic pitch (supra) was also employed in this experiment to scale received pitch. In order to present stimuli with parameter values that sampled systematically a broad range of fundamental and formant frequencies, complex stimuli were prepared that simulated vowel sounds. In view of the parameter values of vowel quality reported by Peterson (1961) and Fant (1960), the following fundamental and first- and second- formant frequencies were employed: F_0, 100, 120, 140, 160, 180, 200; F_1, 250, 350, 450, 550, 650, 750; F_2, 950, 1250, 1550, 1850, 2150, 2450 (cps). The 216 stimuli generated by all combinations of F_0, F_1, and F_2 were tape recorded, in random order, in the following manner. A calibrated oscillator (Hewlett-Packard 200AB) drove a pulse generator that supplied the fundamental frequency and its harmonics to two filters (Krohn-Hite 310 AB) arranged in parallel. The filter outputs, corresponding to the first and second formants, were amplified separately so that F_2 was 6 db less (rms voltage) than F_1. The "formants" were mixed and a 250 msec sample with 100 msec rise time was tape recorded. The stimuli were presented over PDR-8 earphones at approximately 60 db (SPL) to each of 14 Ss for category ratings of vowel pitch.

Echoic Responding

(a) Pitch matching. The synthetic vowel tape, whose preparation is described above, was presented to each of five listeners with instructions to imitate each vowel sound as accurately as possible, "paying particular attention to the pitch of the vowel." Tape recordings were processed to determine the fundamental frequency of the matching response as a function of the F_0, F_1, and F_2 parameter values of the stimuli; pitch measurement procedures are described above. This pitch matching function was predicted from scales of autophonic and received pitch.

(b) Concurrent matching of loudness, duration, and pitch. A schematic of the experimental design appears in Fig. 2; the preparation of the stimulus tape is described above (duration estimation). Nine subjects imitated each of the 64 stimuli in the series, which was presented three times. Each of the 192 x 9 responses was processed to determine its amplitude, duration, and fundamental frequency; measurement procedures are those described above. Each response parameter (averaged across subjects and repetitions) may then be plotted as a function of the relevant stimulus parameter, and this plot may be compared to the predicted matching function based on autophonic and reception scales.

Results and Discussion

Autophonic and reception scales for the speech parameters amplitude, duration, and fundamental frequency are shown in Figs. 3, 4, and 5, respectively. Comparison of these psychophysical functions leads at once to an important generalization: the speaker's perception of his own speech parameter values grows more rapidly as a function of stimulus magnitude than the listener's perception of these same parameters. It will also be observed that the functions relating physical to apparent magnitude for the intensive or "prothetic" (Stevens, 1957) speech parameters, amplitude and duration, are well described by straight lines in log-log coordinates, in other words, a power law. Thus, autophonic and received amplitude and duration may be added to the host of other continua on which psychological magnitude has been demonstrated to be a power function of the stimulus (Stevens, 1960). Because the obtained exponent (slope) of the psychophysical function may be influenced to some extent by the choice of measurement technique, the slopes fit by least squares to the obtained data (Figs. 3, 4) should be taken only as an estimate of those values that will prove most generally descriptive: autophonic amplitude, 1.2; received amplitude, 0.4; autophonic duration, 1.6; received duration, 0.9.

Category estimates of autophonic and received pitch (Fig. 5) are well described by straight lines in semi-log coordinates, in other words, a logarithmic law. As observed for speech amplitutde and duration, estimates of autophonic pitch grow more rapidly as a function of fundamental frequency than estimates of received pitch; the slopes of the functions relating category estimates (arbitrary zero) to log relative pitch are 21 and 11, respectively.

The finding that an exponential equation describes pitch perception is consonant with earlier findings reported by Stevens and Gallanter (1957) in a summary of research on subjective scales of pure tone pitch. Stevens (1957) has suggested that pitch may constitute a metathetic continuum, for which category scales, jnd scales, and ratio scales are linearly related. The category estimates of received pitch, shown as a function of fundamental frequency in Fig. 5, were obtained in an experiment in which first and second formant frequencies were also manipulated systematically. Figure 6 (top) shows that the fundamental frequency of the vowel-like sounds played the major role in determining estimates of pitch but the first and second formants also had appreciable effects. An analogous experiment, in which the subject was instructed to imitate rather than estimate the synthetic vowels, gave comparable results (Fig. 6, bottom). The formant frequencies of the vowel stimuli seem to influence the pitch of echoic responses less, however, than they do numerical estimates of pitch.

A simplified analysis of the mechanics of speech productions suggests that there is an "intrinsic" relation between the pitch and amplitude of the vocal response. It is generally believed that the laryngeal tone is produced by the alternating force exerted on the vocal folds, initially by the subglottal pressure and then, following the spread of the folds, by the negative pressure or Bernoulli effect, caused by the stream of air through the folds. The time for an excursion of the vocal folds may then be related to their distance of travel and the velocity of the air stream through the folds by $T = d\ v^{-1}$. Ignoring the elasticity of the folds and the density of air, $p = v^2$ and $T = d\ p^{-0.5}$. Ladefoged and McKinney (1962) have reported that the peak subglottal pressure is related to the sound pressure produced by $P = S^{0.67}$. Then, $T = dS^{-0.34}$. Since the frequency of vibration of the vocal folds is inversely related to their period, $F = S^{0.34}\ d^{-1}$. In logarithimic coordinates, the change in pitch effected by a change in amplitude is then given by: $\log (F_2/F_1) = .34 \log (S_2/S_1)$. The results of an empirical determination of this relation are shown in Fig. 7 (filled circles). The fundamental frequency of the phoneme /a/, obtained when four autophonic amplitudes were required, grew as the 0.2 power of the sound pressure produced. Within the error introduced by a simplified analysis and the extrapolation of the Ladefoged-McKinney results, we may say that the predicted form and exponent of the pitch-amplitude function has been validated. Figure 7 also shows the effect on response amplitude when five autophonic pitches were required. There is some evidence for an increase in response amplitude over the wide range of pitches employed. Since the subjects were not instructed in how to produce the required amplitudes and pitch levels, it may be that the increase in amplitude at the highest pitch levels reflects S's attempt to "employ" the converse relation described above.

Table I presents parameter values for the vocal response in the experiment involving echoic responding to vowel stimuli that varied in amplitude, duration and pitch. The functions relating stimulus to response parameters may be predicted from the autophonic and reception scales governing parameter perception. In general, if two sensory continua are governed by the equations,

$$\Psi_1 = \Phi_1^m \text{ and } \Psi_2 = \Phi_2^n$$

and if the psychological values, Ψ_1 and Ψ_2, are equated at various levels, it follows that the stimulus values Φ_1 and Φ_2 should stand in the realtion,

$$\log \Phi_1 = (n/m) \log \Phi_2$$

In other words, "cross-modality matches" (Stevens, 1959) should produce a function that is a straight line in log-log coordinates and has a slope given by the ratio of the exponents n and m. In the present experiment, the functions describing echoic responding to stimulus parameters may be employed, therefore, to validate autophonic and reception scales obtained earlier.

Figure 8 shows the relation of response to stimulus amplitude during echoic responding. Since autophonic and reception scales of perceived vowel amplitude (Fig. 3) are power functions of stimulus intensity with exponents 1.2 and 0.4 respectively, the predicted matching function is a straight line in log-log coordinates with slope 0.3. The means of the data points shown in Fig. 8 are well fit by a straight line with the slope 0.33 (method of least squares). It will be observed that the fundamental frequency of the stimulus has large, systematic effects on the amplitude of the matching response. The third stimulus parameter, duration, had a slight and nonsystematic effect on response amplitude (Table I).

The relation between response and stimulus duration is depicted in Fig. 9. Once again, the matching function is well described by the predicted power law. However, the obtained slope of 1.0 is not predicted by the divergent scales of autophonic and received duration. The effects of stimulus amplitude and pitch on response duration are noted in Table I. In general, greater stimulus amplitude or lower pitch yields an increase in response duration. The effects are typically small, however, and never exceed 100 msec.

Although cross-modality matching has, in the past, been employed solely for prothetic continua, the extension to metathetic continua, such as pitch, appears to be straightforward. Since the autophonic and reception scales for vowel pitch are logarithmic functions (Fig. 5), we write:

$$\Psi_A = 21 \log F_A \text{ and } \Psi_R = 11 \log F_R$$

In echoic responding, S is instructed to match: $\Psi_A = \Psi_R$

Then, $21 \log F_A = 11 \log F_R$

and $\log F_A = .52 \log F_R$.

In other words, the predicted pitch matching function is a straight line in log-log coordinates, with slope 0.52. The validity of this prediction may be examined, first, in the context of echoic responding to a single stimulus parameter. Figure 6 shows the pitch of echoic responses to synthetic vowel stimuli as a function of their fundamental frequency. These data were normalized and plotted in logarithmic coordinates in Figure 10 (filled circles; the data have been shifted 0.1 log units along the ordinate for clarity). It is clear that the predicted function, based upon the subjective scales for autophonic and received pitch, provides a close approximation to the obtained matching function.

When the subject was instructed to match vocally the pitch, duration and amplitude of vowel stimuli, (unfilled symbols, Fig. 10), the pitch matching function is essentially the same as in the single-parameter case, but a marked effect of stimulus amplitude may be observed. Vowel duration had no effect on response pitch (Table I). The effect of vowel amplitude on response pitch, which may be examined more readily in Fig. 11, is expected when we recall that S is concurrently matching response amplitude to stimulus amplitude and that autophonic amplitude and pitch covary according to the function $F = A^{0.2}$ (Fig. 7). The relation between log relative stimulus amplitude and response pitch, shown in Fig. 11, is best described by a power function with exponent 0.04. To obtain the relation between response pitch and response amplitude, we note that, in this experiment,

$$\log A_R = .33 \log A_S .$$

Since $\log F_R = .04 \log A_S$,

by substitution, $\log F_R = .12 \log A_R$.

Thus, the relation between response pitch and response amplitude in echoic responding may be predicted reasonably well from the generalized pitch-amplitude function for the vocal response (Fig. 7): $\log F_R = .2 \log A_R$.

Summary

Psychophysical scales are determined for the amplitude, duration and fundamental frequency parameters of vowels. Numerical estimates of vowel loudness and of vowel duration grow as a power function of their respective parameter values, while estimates of pitch are a logarithmic function of fundamental frequency. A given change in the amplitude, duration, or fundamental frequency of a vowel appears greater to the speaker than to the listener; in other words, autophonic scales of vowel parameters grow more rapidly as a function of stimulus magnitude than do reception scales. When a subject is instructed to match his vocal response to a vowel stimulus, autophonic and reception scales

of vowel perception predict the parameters of echoic responding. The matching function for each stimulus parameter shows some influence of the other parameters; the largest effect is an increase in response pitch associated with an increase in stimulus amplitude. This interaction is predicted quantitatively from a simplified analysis of the mechanics of the glottal source and from an empirical determination of pitch-amplitude relations in free responding.

Footnotes

1. The assistance of Mr. D. R. Brinkman is gratefully acknowledged. This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education.

2. For a discussion of circuit design, see Peterson and McKinney (1962).

References

Fant, G. Acoustic theory of speech production. The Hague: Mouton and Co., 1960.

Ladefoged, P. and McKinney, N. P. Loudness, sound pressure, and subglottal pressure in speech. Unpublished report, Communication Sciences, University of Michigan, 1962.

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. I. In: Experimental analysis of the control of speech production and perception. Ann Arbor: University of Michigan report 04411-1-P, 1961.

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. III: Vocal matching of stress patterns. In: Experimental analysis of the control of speech production and perception. Ann Arbor: University of Michigan report 04411-1-P, 1961a.

Lane, H. L., Catania, A. C. and Stevens, S. S. Voice level: autophonic scale, perceived loudness and effects of sidetone. J. Acoust. Soc. Amer., 1961, 33, 160-177.

Lehiste, I. and Peterson, G. E. Vowel amplitude and phonemic stress in American English. J. Acoust. Soc. Amer., 1959, 31, 428-435.

Peterson, G. E. Parameters of vowel quality. J. Sp. Hear. Res., 1961, 4, 10-29.

Peterson, G. E. and McKinney, N. P. The measurement of speech power. Phonetica (in press), 1962.

Stevens, S. S. The measurement of loudness. J. Acoust. Soc. Amer., 1955, 27, 815-829.

Stevens, S. S. On the psychophysical law. Psychol. Rev., 64, 1957, 153-181.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 54, 1958, 177-196.

Stevens, S. S. Cross modality validation of subjective scales for loudness, vibration and electric shock. *J. exp. Psychol.*, 1959, 57, 201-209.

Stevens, S. S. The psychophysics of sensory function. *Amer. Scient.*, 48, 1960, 226-253.

Stevens, S. S. and Galanter, E. H. Ratio scales and category scales for a dozen perceptual continua. *J. exp. Psychol.*, 1957, 54, 377-411.

Stevens, S. S., Volkman, J. and Newman, E. B. A scale for the measurement of the psychological magnitude pitch. *J. Acoust. Soc. Amer.*, 1937, 8, 185-190.

Table I

Parameters of the vocal response in vowel matching. (Each entry is the mean of 30 responses, 3 by each of 10 speakers.)

Response	Amplitude (db re arbitrary zero)				Duration (msec)				Pitch (cps)			
Stimulus	Pitch (cps)				Pitch (cps)				Pitch (cps)			
	220	150	120	100	220	150	120	100	220	150	120	100
Duration (msec)												
					80 db (SPL)							
200	32.4	31.2	28.2	26.2	300	350	250	340	164	154	128	122
300	32.0	29.0	30.2	26.3	410	460	460	500	175	149	133	125
400	32.7	31.0	30.0	26.6	540	590	590	660	164	154	132	125
500	33.2	33.0	29.4	27.1	630	660	680	710	179	161	128	123
					70 db (SPL)							
200	27.7	25.4	24.6	21.0	280	330	190	320	156	152	122	114
300	26.2	24.9	25.0	23.8	340	390	390	430	154	147	127	119
400	27.3	27.3	24.3	22.5	510	550	570	610	161	143	123	114
500	30.1	27.5	25.0	21.8	570	640	650	660	161	147	123	115
					60 db (SPL)							
200	24.5	21.9	20.3	18.3	230	270	170	240	137	145	118	114
300	25.0	23.3	22.3	17.6	350	350	380	410	169	143	127	105
400	25.3	25.7	22.3	18.2	540	490	530	490	156	143	125	110
500	25.2	24.5	22.4	18.5	610	630	660	600	152	147	125	111
					50 db (SPL)							
200	21.9	23.0	19.9	16.0	220	260	200	240	145	149	127	106
300	22.0	19.9	19.6	18.2	300	330	370	360	149	135	119	108
400	21.7	21.8	20.2	16.9	460	480	520	480	152	147	119	103
500	18.1	20.1	18.2	14.8	580	590	560	500	135	137	115	109

Figure Captions

Fig. 1. Vowel parameters and their perceptual correlates that determine parameters of an echoic response. A_S, D_S, P_S represent the amplitude, duration, and fundamental frequency parameter values for the stimulus, while A_R, D_R, and P_R represent those for the autophonic response. The subscript "p" means perceived magnitude.

Fig. 2. Schematic representation of an experimental design used to study the relations among vowel parameters and vowel perception.

Fig. 3. Autophonic and reception scales for vowel amplitude. The magnitude production data, shown by filled circles, were taken from Lane et al. (1961); each point represents the logarithm of the relative mean amplitude of three responses by each of 24 subjects. The magnitude estimation data, unfilled circles, were taken from Lane (1961); each point represents the logarithm of the relative mean of four estimates by each of 10 subjects.

Fig. 4. Autophonic and reception scales for vowel duration. Each filled circle is the logarithm of the relative mean duration of 12 magnitude productions by each of 10 subjects. Each unfilled circle is the logarithm of the relative mean of 48 magnitude estimations by each of 10 subjects.

Fig. 5. Autophonic and reception scales for vowel pitch. Each filled circle is the mean category estimate of five autophonic pitches produced by each of nine subjects. Each unfilled circle is the mean category estimate of 36 vowel pitches received by each of 14 subjects.

Fig. 6. The effect of the fundamental and first- and second-formant frequencies on the perceived pitch of vowel-like sounds. The means of 216 pitch estimates by each of 14 subjects (top) and the mean pitch of 216 echoic responses by each of five subjects (bottom) are plotted as a function of the fundamental frequency (F_0), first formant frequency (F_1) and second formant frequency (F_2) of the stimuli.

Fig. 7. The relation of autophonic pitch to autophonic amplitude. Filled circles: each of ten subjects was required to produce four autophonic levels of the vowel /a/ three times, in irregular order. The mean fundamental frequency of the responses at each amplitude was divided by that parameter value at the lowest amplitude, and the logarithm of this relative mean pitch was plotted. Unfilled circles: each of ten subjects was required to produce five pitch levels of the vowel /a/, three times in irregular order. The mean peak amplitude of the responses at each pitch level was divided by that parameter value at the lowest pitch level, and the logarithm of this relative mean amplitude was plotted.

Fig. 8. Echoic responding to vowels: the effect of stimulus amplitude and pitch on response amplitude. Each point is the mean decibel level of 12 responses by each of nine subjects.

Fig. 9. Echoic responding to vowels: the effect of stimulus duration on response duration. Each point is the logarithm of the relative mean duration of 48 responses by each of nine subjects.

Fig. 10. Echoic responding to vowels: the effect of stimulus pitch and

amplitude on response pitch. Filled circles: each point is the logarithm of the relative mean pitch of 36 echoic responses by each of 14 subjects, matching synthetic vowels with constant amplitude and duration. The matching function predicted from autophonic and reception scales of vowel pitch is shown. Unfilled symbols: each point is the logarithm of the relative mean pitch of 12 echoic responses by each of nine subjects, matching three vowel parameters concurrently.

Fig. 11. Echoic responding to vowels: the relation between stimulus amplitude and response pitch during concurrent matching of three vowel parameters. Each point is the logarithm of the relative mean pitch of 12 echoic responses by each of nine subjects.

$$\boxed{\begin{array}{ll} & \left[\begin{array}{l} \underline{A}_{S_P} = g(\underline{A}_S, \underline{D}_S, \underline{P}_S) \\ \underline{D}_{S_P} = h(\underline{D}_S, \underline{A}_S, \underline{P}_S) \\ \underline{P}_{S_P} = i(\underline{P}_S, \underline{A}_S, \underline{D}_S) \\ \underline{A}_{R_P} = j(\underline{A}_R, P_R, D_R) \\ \underline{P}_{R_P} = k(\underline{P}_R, A_R, D_R) \\ \underline{D}_{R_P} = l(\underline{D}_R, A_R, P_R) \end{array}\right] \\ \begin{array}{l} \underline{A}_R = f_1 \\ \underline{D}_R = f_2 \\ \underline{P}_R = f_3 \end{array} & \end{array}}$$

Fig. 1

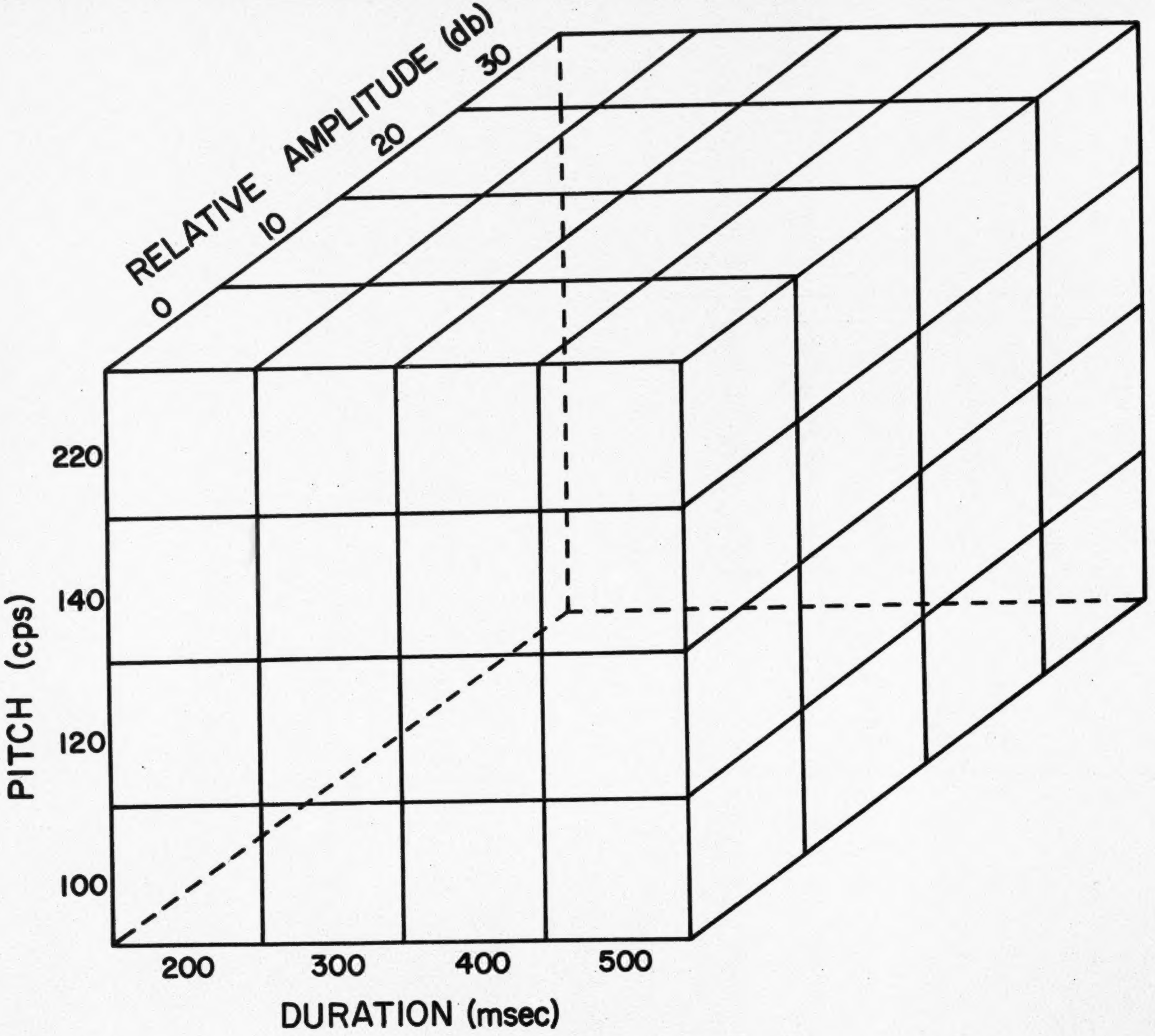
RELATIVE AMPLITUDE (db)
PITCH (cps)
DURATION (msec)

Fig. 2

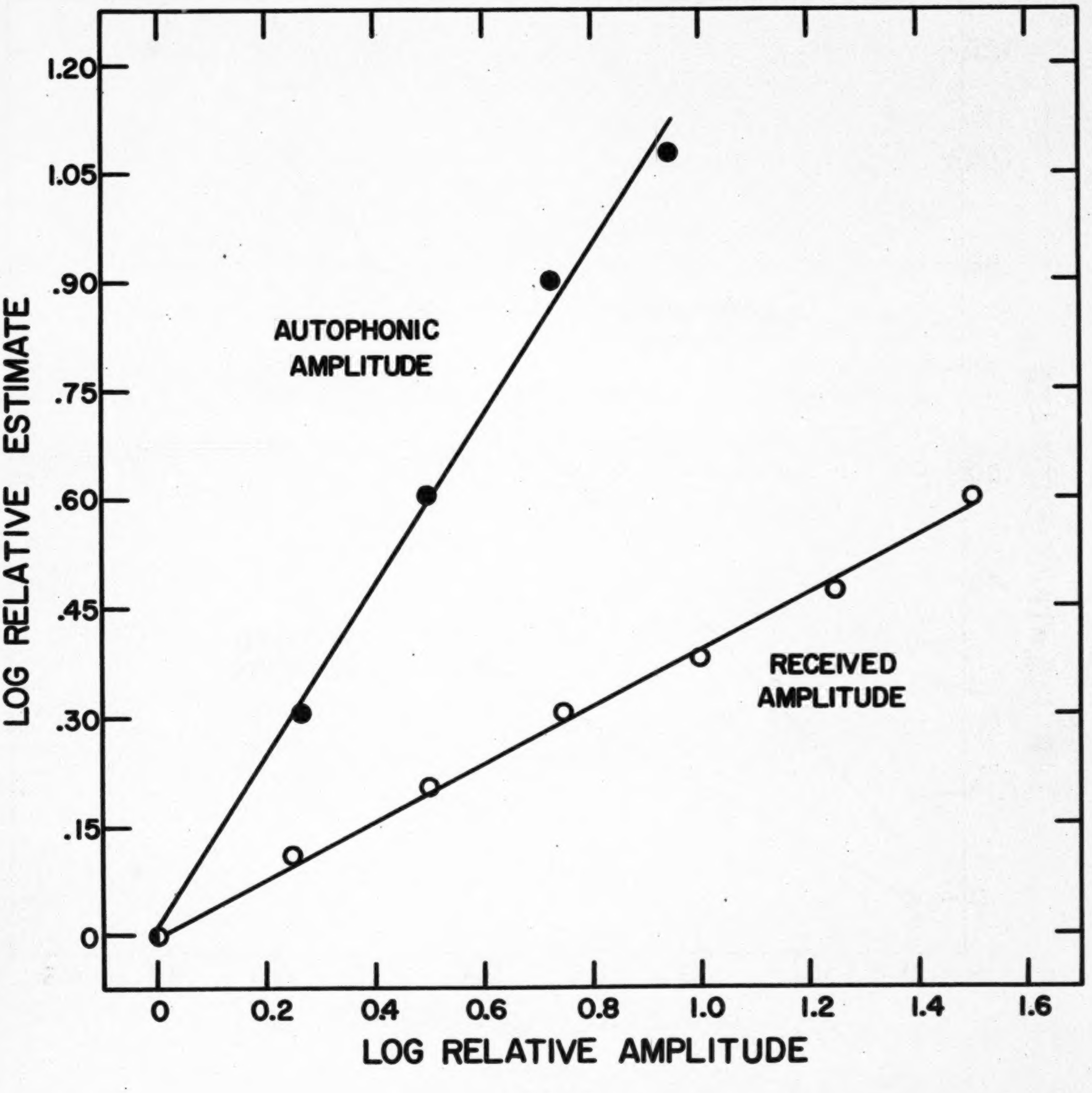
1.20
1.05
.90
.75
.60
.45
.30
.15
0
LOG RELATIVE ESTIMATE
AUTOPHONIC AMPLITUDE
RECEIVED AMPLITUDE
0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
LOG RELATIVE AMPLITUDE

Fig. 3

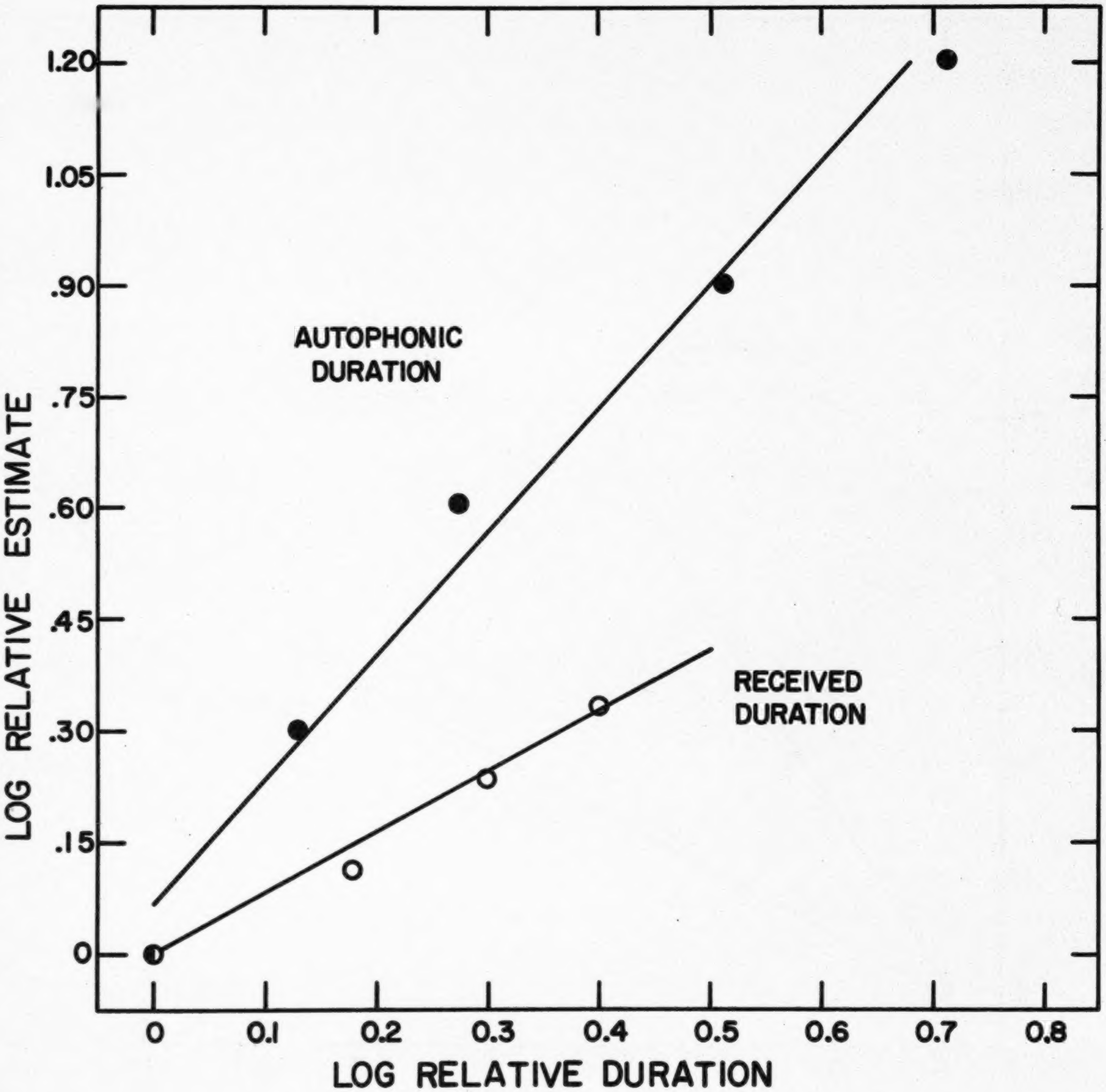
1.20
1.05
.90
.75
.60
.45
.30
.15
0
LOG RELATIVE ESTIMATE
AUTOPHONIC DURATION
RECEIVED DURATION
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
LOG RELATIVE DURATION

Fig. 4

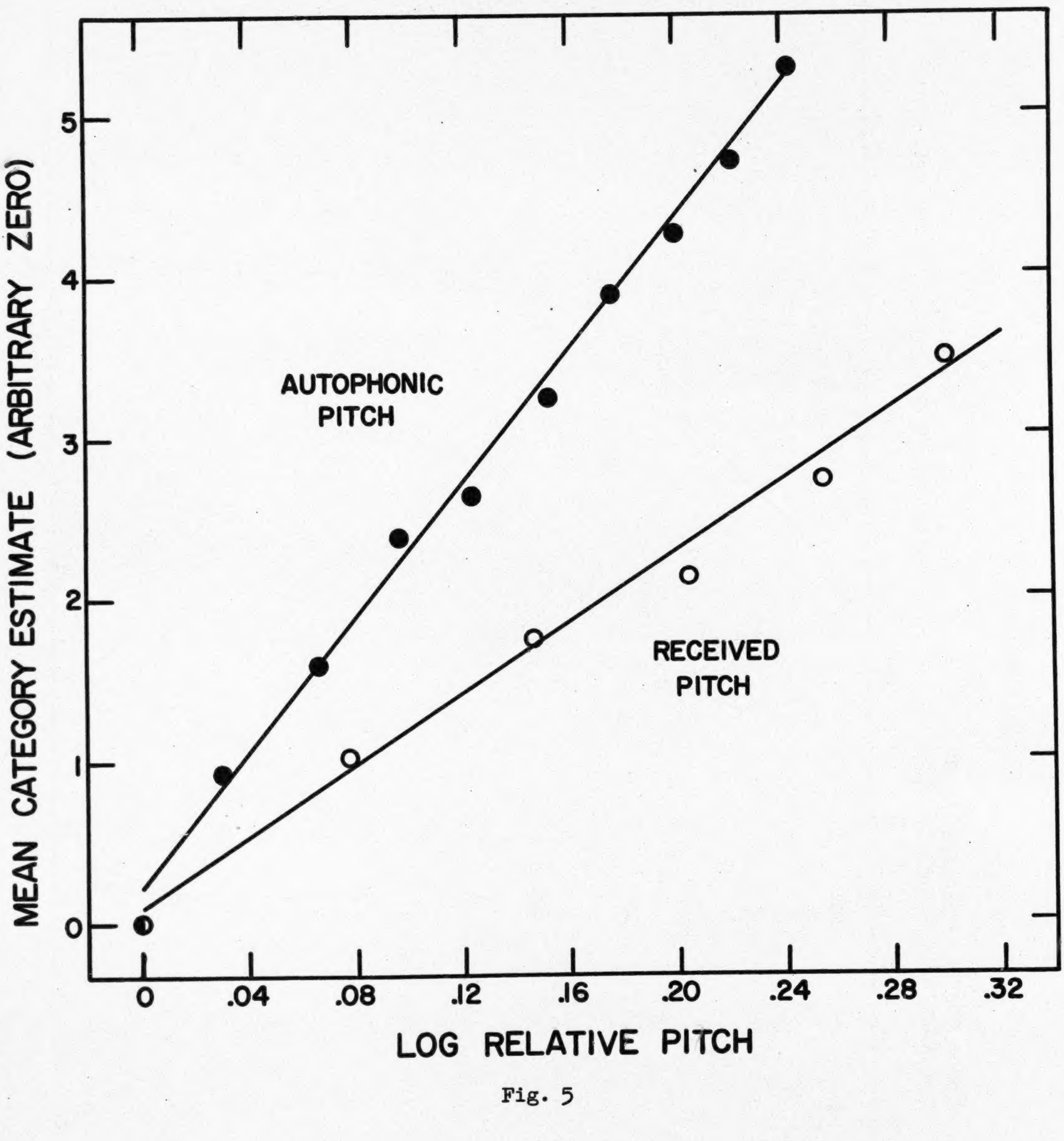
MEAN CATEGORY ESTIMATE (ARBITRARY ZERO)
0
1
2
3
4
5
0
.04
.08
.12
.16
.20
.24
.28
.32
LOG RELATIVE PITCH
AUTOPHONIC PITCH
RECEIVED PITCH

Fig. 5

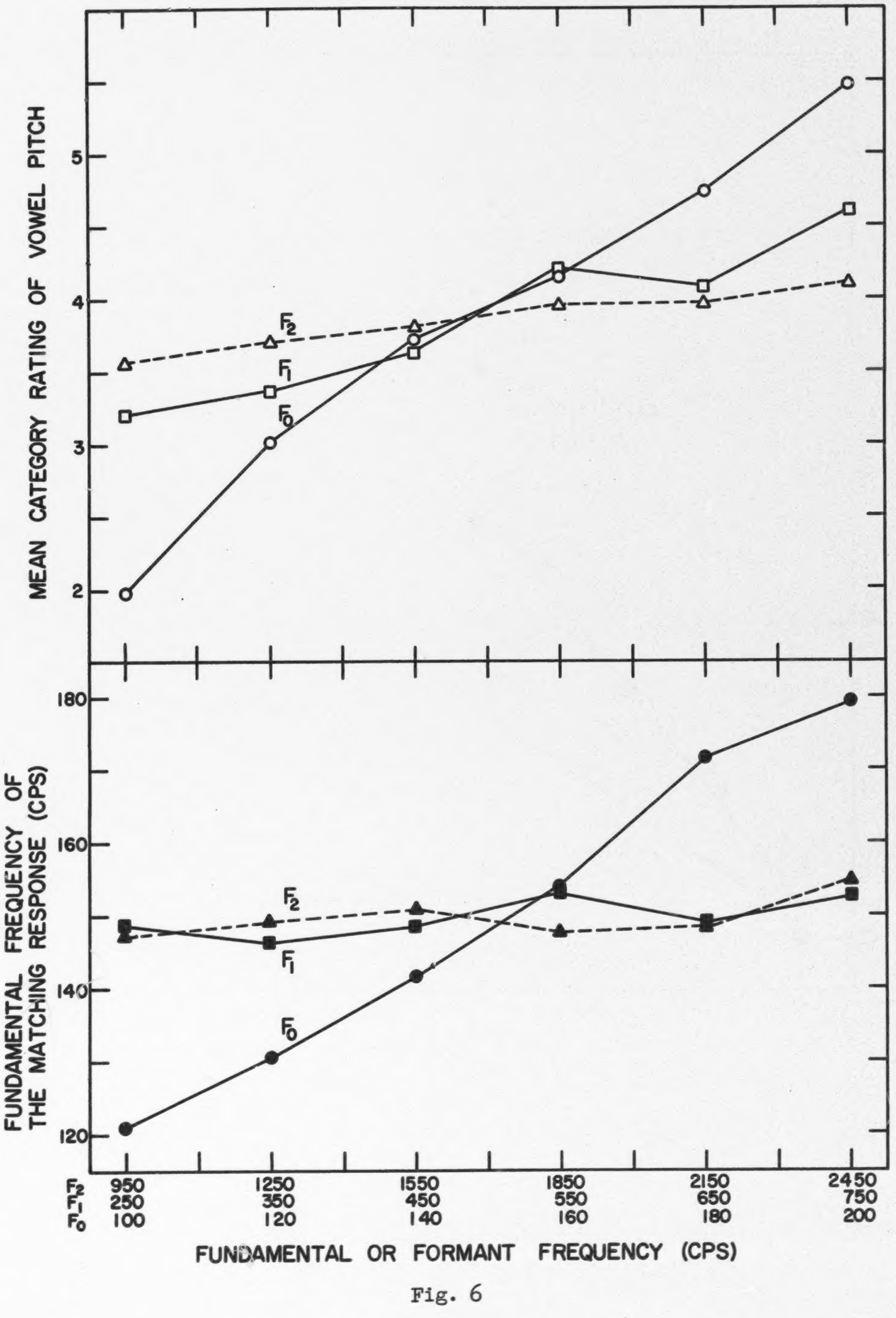
MEAN CATEGORY RATING OF VOWEL PITCH
2
3
4
5
F_2
F_1
F_0
FUNDAMENTAL FREQUENCY OF THE MATCHING RESPONSE (CPS)
120
140
160
180
F_2
F_1
F_0
F_2 950 1250 1550 1850 2150 2450
F_1 250 350 450 550 650 750
F_0 100 120 140 160 180 200
FUNDAMENTAL OR FORMANT FREQUENCY (CPS)

Fig. 6

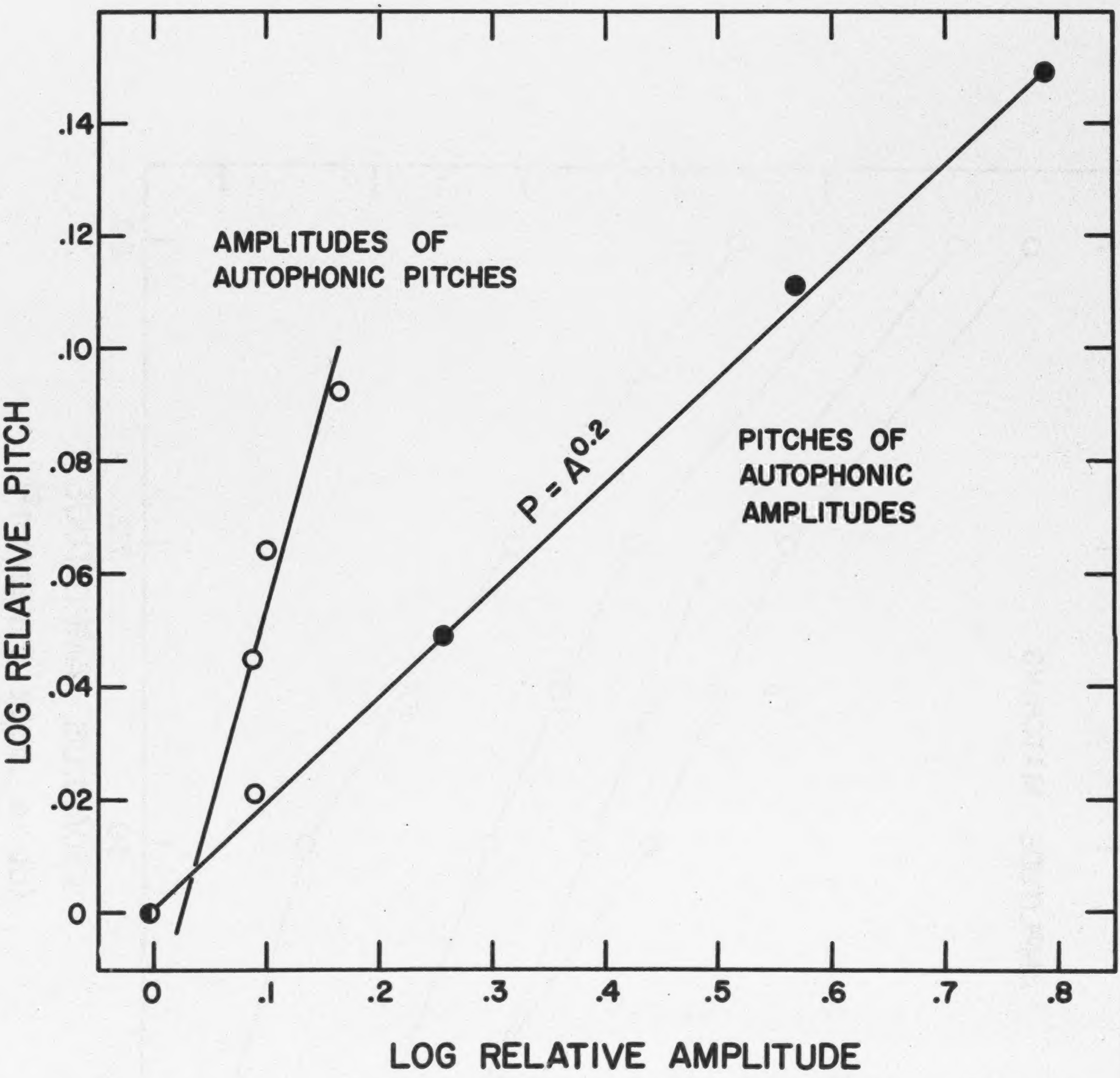
.14
.12
.10
.08
.06
.04
.02
0
LOG RELATIVE PITCH
0
.1
.2
.3
.4
.5
.6
.7
.8
LOG RELATIVE AMPLITUDE
AMPLITUDES OF AUTOPHONIC PITCHES
PITCHES OF AUTOPHONIC AMPLITUDES
P = A0.2

Fig. 7

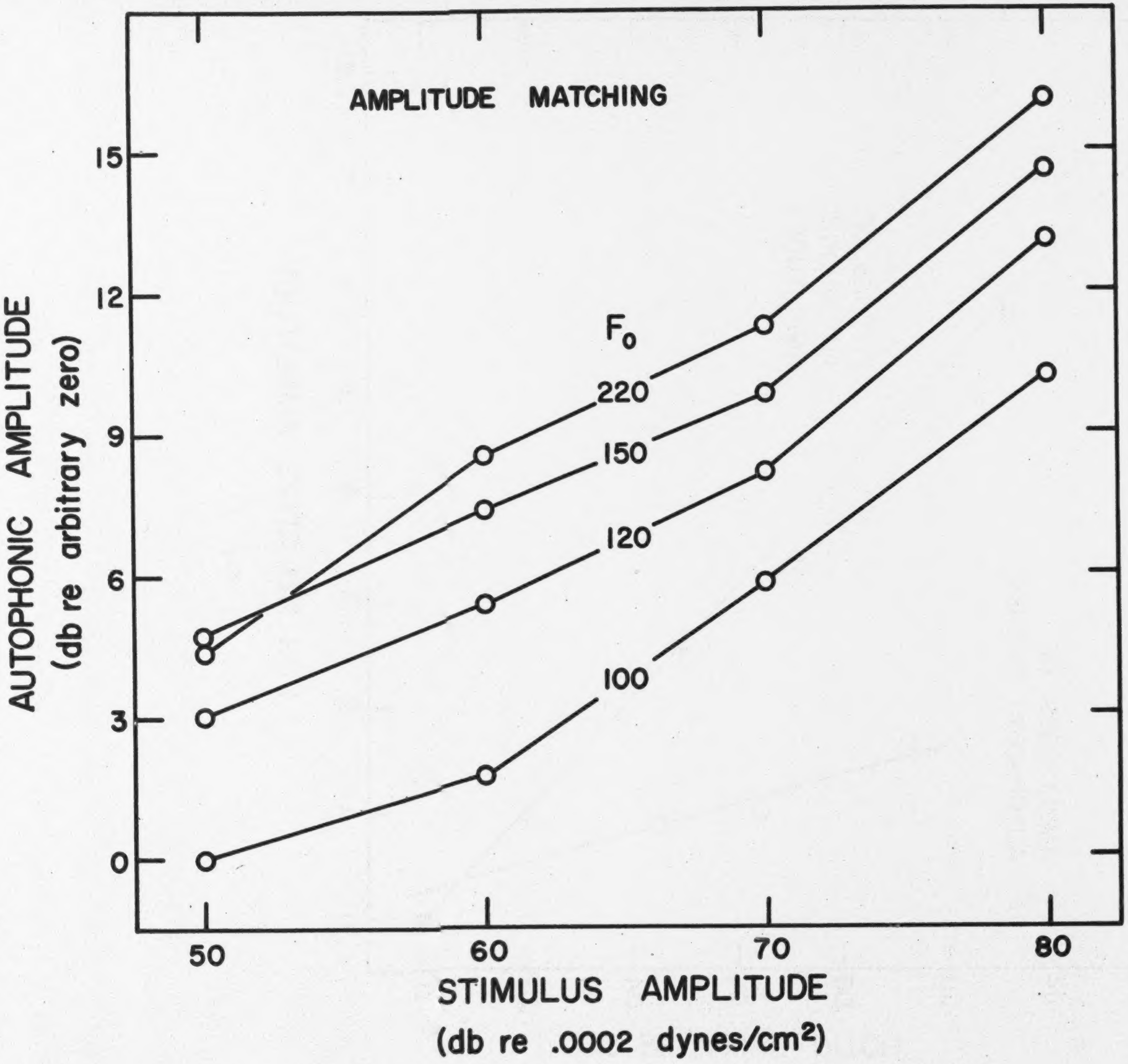
AMPLITUDE MATCHING
F$_0$
220
150
120
100
AUTOPHONIC AMPLITUDE
(db re arbitrary zero)
0
3
6
9
12
15
50
60
70
80
STIMULUS AMPLITUDE
(db re .0002 dynes/cm^2)

Fig. 8

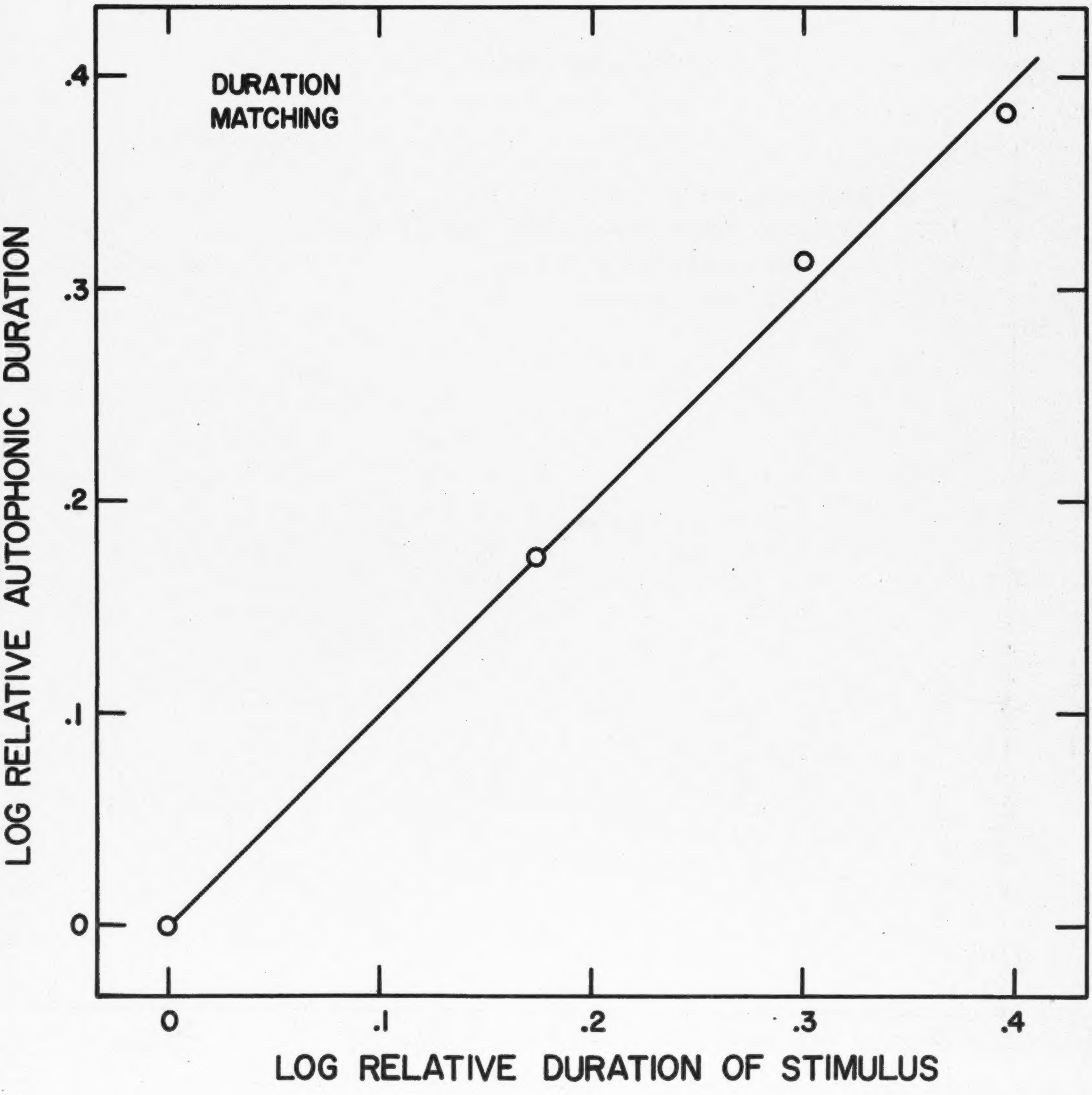
DURATION
MATCHING
LOG RELATIVE AUTOPHONIC DURATION
0
.1
.2
.3
.4
LOG RELATIVE DURATION OF STIMULUS
0
.1
.2
.3
.4

Fig. 9

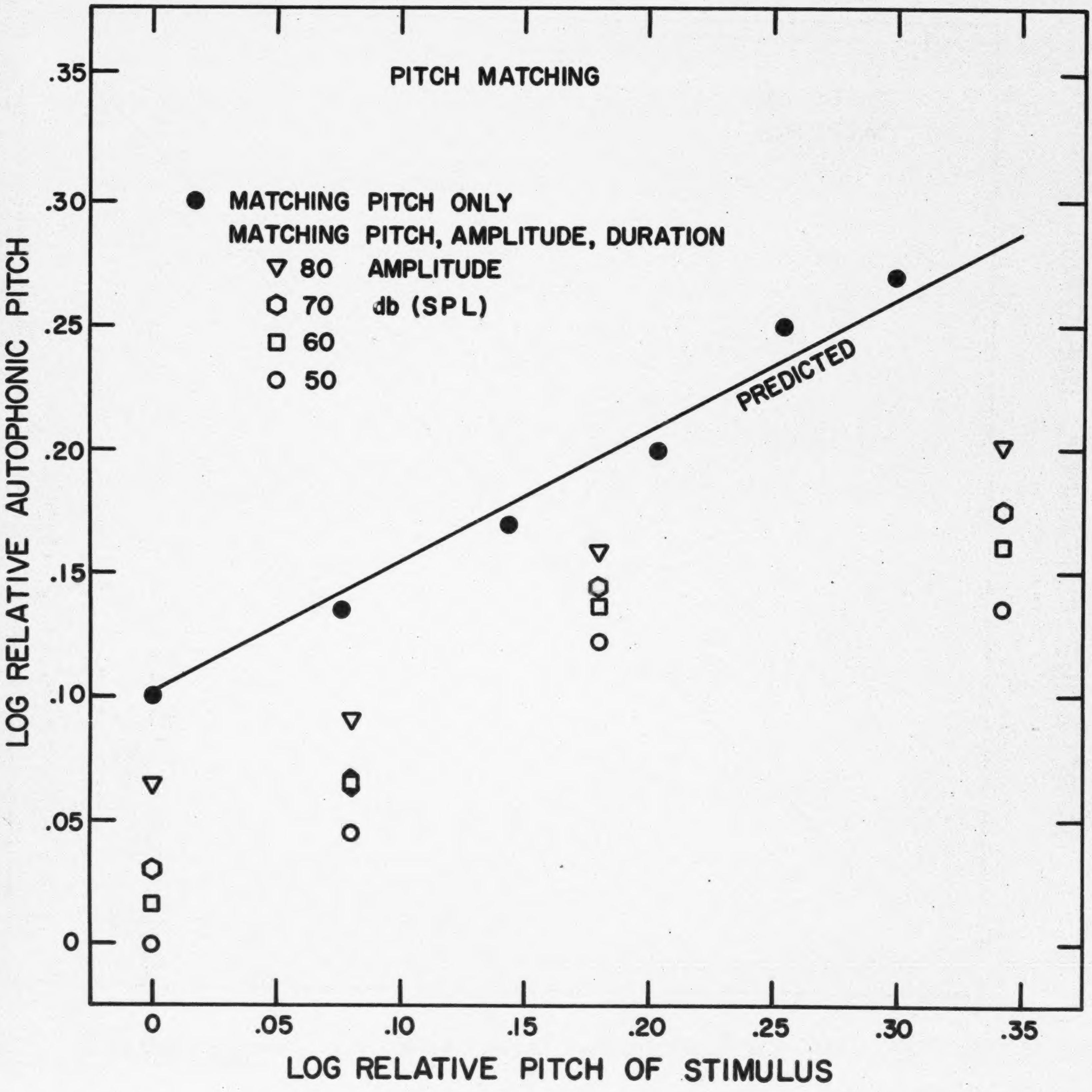
PITCH MATCHING
● MATCHING PITCH ONLY
MATCHING PITCH, AMPLITUDE, DURATION
▽ 80 AMPLITUDE
⬡ 70 db (SPL)
□ 60
○ 50
PREDICTED
LOG RELATIVE AUTOPHONIC PITCH
.35
.30
.25
.20
.15
.10
.05
0
LOG RELATIVE PITCH OF STIMULUS
0
.05
.10
.15
.20
.25
.30
.35

Fig. 10

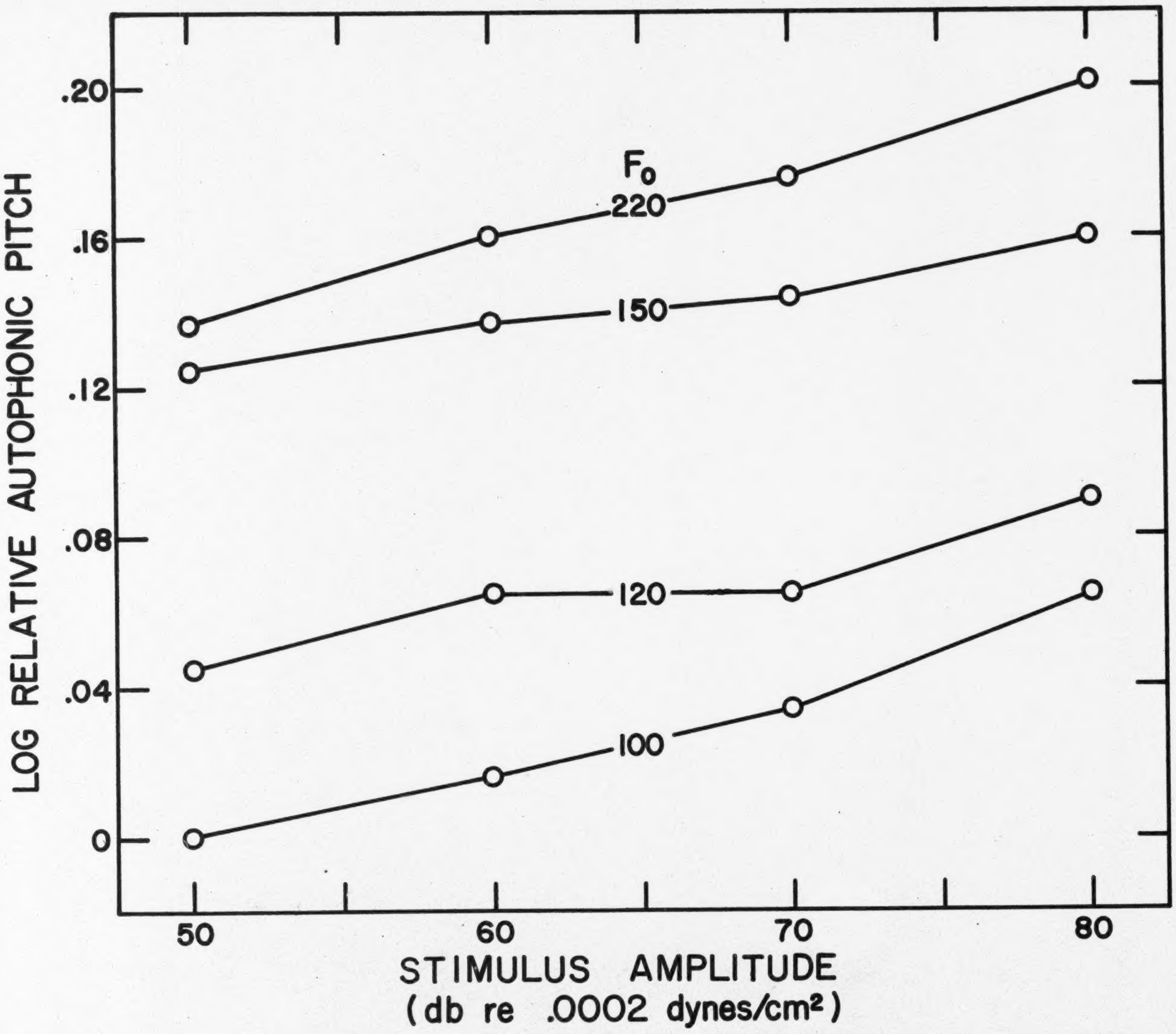
F$_0$
220
150
120
100
.20
.16
.12
.08
.04
0
50
60
70
80
LOG RELATIVE AUTOPHONIC PITCH
STIMULUS AMPLITUDE
(db re .0002 dynes/cm²)

Fig. 11

THE RELATIONS AMONG RESPONSE RATE, TOPOGRAPHY, AND SCHEDULES OF REINFORCEMENT: METHODS AND FINDINGS IN AN ANALYSIS OF THE VOCAL OPERANT

Harlan Lane and P. G. Shinkman
The University of Michigan

The relations among response rate, topography, and schedules of reinforcement: methods and findings in an analysis of the vocal operant

Harlan Lane and P. G. Shinkman[1]
Communication Sciences Laboratory
University of Michigan[2]

Lately, there has been a considerable increase in research in two heretofore unrelated areas: the rate of emission of vocal operants (Flanagan et al., 1958; Lane, 1960, 1961; Shearn et al., 1961; Starkweather, 1960; Starkweather and Langsley, 1961) and the topographical properties of non-vocal behavior (Goldberg, 1959; Margulies, 1961; Millenson et al., 1961; Notterman, 1959; Schaefer and Steinhorst, 1959)[3]. It has been established, in the first area cited, that the rates of emission of human and infra-human vocal responses are amenable to reinforcement control. The second area of research has shown that there are systematic functional relations among response topography, response rate, and contingencies of reinforcement. It has not been known to what extent these relations apply to vocal behavior.

There is a certain irony, therefore, in observing that the vocal response may be preferred, on several counts, to any other operant for an inquiry into the topographical properties of operant behavior in general. Unlike many other operants, whose muscular constituents are relatively inaccessible for measurement, "the complex muscular responses of vocal behavior affect the verbal environment by producing audible 'speech'. This is a much more accessible datum" (Skinner, 1957). The usefulness of this datum is predicated on the substantial evidence (Fant, 1960) that changes in the complex muscular responses of vocal behavior are closely correlated with changes in the less numerous acoustic parameters of speech. The facility of measuring the vocal response is enhanced by the availability of advanced instrumentation for the storage and measurement of acoustic signals. An interest in vocal behavior within related disciplines has lead, moreover, to a degree of sophistication in parameter portrayal which surpasses that for any other operant.

Aside from the role that vocal topography may play as a vehicle for a more general inquiry, it would seem to warrant research in its own right. In the prediction and control of vocal behavior, we often must deal with a single instance of the operant. In this case, frequency of emission cannot be employed as an index of "response strength" and interest centers upon such topographical or intensive properties of the response as amplitude, pitch, and duration.

A consideration of methods for measuring vocal topography and an inquiry into its relation, on the one hand, to vocal rate and, on the other, to non-vocal topography would seem to be in order.

Method

The relation of measurement parameters to vocal topography. It was not desirable in the present research to define the response under study as the closure of a voice-operated relay when suitably juxtaposed with a subject, because interest centered on the subclasses of this operant. Alternatively, the response could have been defined as some voltage function of time, but a more general and useful definition would be in terms of the acoustic signal radiated from the subject's mouth. This vocal response, or acoustic event, may be represented by a large number of parameters; the following were selected in the present study: peak average amplitude, mean duration of the initial ten periods of the fundamental frequency, and duration of the total acoustic event. These parameters are neither exhaustive nor independent. Although the choice of response parameters may be determined eventually by the degree to which they show systematic relation to experimental variables, it was made initially in view of their relation to the speech production mechanism. To a large extent, the speech pressure function of time is predictable from the configuration of the vocal source and tract; the converse is also true (Delattre, 1951).

(1) Peak average amplitude. It is generally thought that the laryngeal tone is produced by the alternating force exerted on the vocal folds, initially by the subglottal pressure and then, following the spread of the folds, by the negative pressure or Bernoulli effect, caused by the stream of air through the folds. The major determinant of the peak average amplitude (hereafter called amplitude) is the peak subglottal pressure which, in man, is caused by the contraction of the respiratory muscles. Ladefoged and McKinney (1962) report that the sound pressure level of sustained vowels is proportionate to the corresponding subglottal pressure to the 1.5 power.

The pulsating airflow through the glottis is a saw-tooth shaped periodic time function which can be expressed as a harmonic spectrum, according to the Fourier transform. The vocal tract above the glottis may be considered a variable filter system. Its effects are represented by multiplying the amplitude of each harmonic of the source spectrum by the gain factor of the filter function at each frequency. The resultant spectrum envelope is the Fourier transform of the pressure function of time radiated from S's mouth. It should be clear, therefore, that the amplitude of the vocal response is determined not only by the subglottal pressure but also by changes in the spectrum of the glottal waveform such as would be caused by a change in the elasticity of the vocal folds, and by changes in the vocal tract filter system, caused by a change in the positions of the articulators. Thus, pitch and vowel quality also influence amplitude.

In order to measure the amplitude of the vocal response, the pressure function of time was transduced by a calibrated dynamic microphone (Altec 633A) and tape recorded (Ampex 300). The recorded signal was applied to an intensity meter[4] that, first of all, introduced full-wave linear rectification. Such a signal has a high information rate which must be reduced for parametric representation; typically, a filter with bandwidth less than that of the original wave is employed. The cutoff frequency must be low enough to attenuate the ripple components due to the quasi-periodic voice source, but high enough to permit accurate measurements of speech transients. A bandwidth of 32 cps was employed (11 msec integrating time) for human voice measurements and one of 150 cps (2 msec integrating time) for chick voice measurements. The meter had linear amplitude compression and no pre-filtering. Its output was applied to a d-c amplifier with flat frequency response (Krohn-Hite DCA-10R) and thence to a peak reading voltmeter (Control

Devices PTM-7-R). This device read and stored the peak of the average amplitude function of time and applied a proportionate voltage to a d-c VTVM (Hewlett-Packard 405CR), which, in turn, encoded the impressed signal and transferred the information to a printout counter (Hewlett-Packard 560AR). The counter recorded the voltage (to three digits) and cleared the peak meter. It was not necessary for the present study to calibrate the measurement system with respect to sound pressure levels at the source. The units of response amplitude are arbitrary, therefore, and only relative values are considered.

(2) Fundamental frequency. Changes in the fundamental frequency of the voice (often called vocal pitch) are determined primarily by the degree of contraction of the thryo-aretenoid muscles, which regulate the elasticity of the glottal margins, and secondarily by the subglottal pressure.

A simplified analysis of the mechanics of the sound source described above suggests that the period of the laryngeal tone (or excursion of the vocal folds) is inversely proportional to the square root of the subglottal pressure, in the absence of any compensatory adjustment of the vocal folds. A psychophysical determination of this relation showed that the relative frequency was proportional to the 0.2 power of the relative amplitude (Lane, 1962). These two parameters of the vocal response may normally be expected to covary, therefore. The fundamental frequency may also be influenced by major constrictions in the vocal tract; this is not of concern in the present study where the vocal response was a vowel.

In human phonation, the fundamental frequency may be considered a population parameter inferred from a distribution of sample period durations of the laryngeal tone; this because of the quasi-periodic vibration of the vocal folds. In chick phonation, the concept of a fundamental frequency is particularly inappropriate since the period of vibration of the tympanic membranes in the syrinx is constantly changing (see Fig. 1). Although the term fundamental frequency is used in the present study, it should be understood that the mean duration of the initial ten periods was measured and then converted to cycles per second. The fundamental frequency was selected from the complex speech wave by applying the tape-recorded signal to two band-pass filters in series (Krohn-Hite 310 AB; 48 db/octave). The filter settings were determined initially by spectrographic analysis (Western Electric BTL 2) and then adjusted to provide better than 30 db rejection of the first harmonic. The filtered signal was sent to the "ten period average" circuit of a frequency counter (Hewlett-Packard 522 B). The mean period was read in milliseconds to two decimal places and recorded by a parallel printer (Hewlett-Packard 560 A). Figure 1 shows that the ten period average reflects the fundamental frequency reasonably well for human vocal responses during CRF but poorly under the other conditions of the experiment. Information reduction was bought at the cost of omitting other marked changes in the topography of the vocal response.

(3) Duration. The duration of the vocal response was measured from the tape-recorded signal by a calibrated voice-operated relay (Miratel). This device consists essentially of an amplifier, full wave rectifier, peak regulator, and a relay. Although the circuit design of this and other VOR's incorporates a peak regulator to provide

drop-out time independent of signal level, this condition holds only for waveforms of relatively rapid rise-decay times. In the voicing of isolated vowels, the rise time may exceed ten per cent of the vowel duration. The measured duration of a signal with a triangular waveform will depend on the relation of signal amplitude to VOR threshold. If signal duration is to be measured independent of amplitude fluctuations, input signals must be processed with a fast-acting automatic volume control, or extensively peak clipped, or some equivalent operation. In the present study, the input signals were amplified to near maximum input level and the VOR threshold level was set 33 db below. When the VOR was operated, the relayed applied a d-c voltage to the trigger circuit of the time interval section of a frequency counter (Hewlett-Packard 522B). The duration was read in milliseconds by the counter and recorded by the parallel printer.

(4) Rate of responding. A voice-operated relay was also employed to provide a cumulative record of the number of responses as a function of time. These data were collected during the experimental session by applying the transduced acoustic signal to the VOR. This device triggered a monostable multivibrator that provided a pulse of fixed duration to a Gerbrand's cumulative recorder. It is important to note that the VOR threshold was set sufficiently low to respond to all voiced signals. Examination of concurrent tape recordings revealed that "false counts" due to non-vowel sounds (coughs and the like) were sufficiently infrequent.

Subjects. Ss were two male and seven female University of Michigan undergraduates and one chick.

Apparatus and procedure.

(1) Human. A modification of Holland's procedure (1957) for the study of observing behavior was used. S sat in a sound-insulated chamber, facing a dynamic microphone (Altec, 633A), with her head fastened in a headrest, to insure that the distance from mouth to microphone remained constant throughout the experiment. A loudspeaker was located behind S's chair. Pencil and paper were presented and the following instructions read:

> "This is an experiment in speech. You will hear numbers read to you over the loudspeaker in groups of about five or six. Each time a group of numbers is read, your job is to write down the numbers in a row of cells on your response sheet. Start a new row for every group of numbers. Numbers are presented only when you say /u/ into the microphone in front of you. Try not to make any other sounds at all, as this may disturb the experiment. The object is to see how many numbers you are able to write down correctly during the experiment, which will last about three [two] hours. Try and stay in the position the experimenter puts you in, throughout the experiment. Are there any questions? The experiment will begin a few seconds after I leave the room."

(Questions were answered by repeating the instructions.)

Recording and control apparatus were located in an adjacent room and arranged in the following way. The subject's vocal responses were transduced, sent to a tape recorder (Ampex 300), and also to a voice-operated relay (Miratel) governing the reinforcement circuit. A second tape recorder (Uher III), which ran continuously during the experiment, contained a tape on which random numbers had been recorded at intervals of about one second. Reinforcement occured when a DPDT electronic switch (Grason-Stadler 821S119) closed for 6.25 seconds, allowing the Uher output to reach the loudspeaker located behind S and, at the same time, disconnecting the microphone in front of S.

There were two experimental conditions. In the first, which lasted about three hours, S was given 15 minutes of continuous reinforcement (CRF), followed by 40 reinforcements on a variable-interval schedule (VI), followed by 73 minutes of extinction (EXT). In the VI schedule of reinforcement there were eight intervals each of 16, 32, 64, 128, 256 seconds in random order. The second experimental condition, which lasted about two hours, was identical to the first, except that CRF was extended beyond the initial 15 minutes to include 40 additional reinforcements, which replaced the VI reinforcements of condition 1.

Six Ss served in condition 1 (CRF-VI-EXT) of the experiment and three Ss in condition 2 (CRF-CRF-EXT). An analysis of response topography was performed for two Ss from each group.

(2) Chick. The experimental procedure for the month-old Bantam chick was comparable to that for the human subjects. No instructions were necessary, however; a convenient vocal response was observed to have a high operant level (2/sec) in the experimental space (modified pigeon chamber). To eliminate the noise generated by pecking, etc., which triggered the VOR, most of the surfaces of the chamber were covered with a tough-skinned foam rubber while the remainder, including

the food bin and water cup, were coated with room-temperature vulcanizing rubber. A light source and photocell were arranged opposite each other in the walls of the food bin and a dynamic microphone was placed just above the bin, adjacent to the magazine light. The chick was conditioned with food reinforcement to hold his head in the bin, thus interrupting the light beam, during the course of the experiment. It was possible, therefore, to arrange and monitor that the chick's head was in a narrowly defined region around the microphone, and to reduce in this way an artifact in amplitude measurements of the vocal response.

The chick was 24 hours food-deprived and at about 80 per cent of free feeding weight, determined five days before, when placed in the chamber on each of three experimental days. On day 1, 15 minutes of CRF was programmed: each vocal response produced four seconds of food reinforcement (Wirthmore Chick Crumbles). Responses occuring during the reinforcement cycle, however, had no effect. On day 2, the same VI schedule of reinforcement employed with the human subjects was programmed for the chick. On day 3, four VI reinforcements were presented and then extinction was in effect for one hour. Cumulative records and tape recordings were collected during the experimental sessions and analyzed subsequently in the manner described above.

Results and Discussion

The cumulative records obtained from nine human Ss and one chick under the two reinforcement sequences (CRF-VI-EXT and CRF-CRF-EXT) are shown in Figs. 2 through 5. The rates of vocal responding by the four

Ss whose data were selected for an analysis of response topography are presented in Fig. 2. In the first VI interval, comparable to the first 256 secs of EXT for Ss 3 and 4, all Ss show a rapid decline in rate with about the same time course. Following the first reinforcement under the VI schedule, there is a rapid local, and gradual overall, increase in the rate of responding by Ss 1 and 2. These Ss received as many reinforcements prior to EXT as Ss 3 and 4 but their history of VI conditioning lead to a much higher rate of responding in EXT.

Inference from the rates of responding observed for Ss 1 and 2 when the session was terminated suggests that a number of additional responses would have been observed with prolonged extinction. The extinction session was considerably prolonged for S5 (Fig. 3), contrary to the instructions that were read. The subject was left undisturbed in the closed audiometric room from the beginning of CRF, when his head was taped to the headrest, until 13 hours later, when the session was terminated and the tape removed. Following 117 reinforcements in CRF and 60 reinforcements in VI, S5 emitted over 8,000 responses in 11 hours of extinction (Table I). This considerable "resistance to extinction" is characteristic of operant behavior following VI conditioning. The time course of extinction for S5 is similar to that obtained from pigeons in extinction after VI conditioning (Ferster and Skinner, 1957, p. 348 ff). Figure 4 presents cumulative records obtained under the two experimental conditions for four additional human Ss. The data for S6 and S7 are comparable to those reported earlier: S8 shows a particularly low rate during VI reinforcement and a correspondingly low rate during EXT. The cumulative record for S9 has an unexplained discontinuity after 20 minutes of extinction. The rates of vocal responding obtained from the chick (Fig. 5) under CRF, VI and EXT schedules of reinforcement have the same relative properties as those of the human Ss but are higher overall (Table I).

Three parameters of the topography of the vocal response were recorded concurrently and analyzed for four human subjects and one chick. Table I presents the mean and variance of the amplitude, duration and fundamental frequency of reinforced responses in CRF and VI and unreinforced responses in EXT. These data lose their dimensionality but are more comparable when normalized with respect to parameter values under CRF. Table II presents the ratio of the parameter values obtained from each S under VI, CRF and EXT to the corresponding values obtained during the initial 15 minutes of CRF.

An important relation between the schedules of reinforcement employed and their effects on response topography is immediately apparent from Table II. All the parameter ratios in Column I are appreciably greater than 1.0. This means that the mean and variance of all three parameters are greater for reinforced responses in VI than for reinforced responses in CRF. Comparable findings have been obtained by Millenson et al. (1961) for the duration of bar-press in the rat during periodic reinforcement: "When rats are exposed to FI contingencies following CRF, unreinforced responses are emitted, and the central tendency and dispersion of the durations of these responses remains two to three times higher than the corresponding values under CRF." These findings confirm the analysis presented by Goldberg (1959) in a discussion of the relation of response variability to resistance to extinction: "In periodic or a-periodic reinforcement, however,... each response will be followed by a period during which responses subsequent to the reinforced one will not be rewarded. This extinction of responses will have the consequence of decreasing the probability of the emission of response forms similar to each previously reinforced one. The resultant number of response forms which will be available for periodic reinforcement will be expanded and the variability of those responses which are periodically reinforced will be greater than the variability of regularly reinforced responses."

The prediction of greater variability among reinforced responses in VI than in CRF, confirmed by the present findings, is predicated on the assumption and considerable evidence that variability increases in extinction. Antonitis (1950) has shown that variability in the locus of a nose-insertion response by the rat increases in EXT. Increased variability in EXT may also be inferred from the following descriptions of bar-press data reported by Skinner (1938): "Stronger responses generally occur near the beginning of the extinction and give way to an unusually low force which is then steadily maintained." "When reinforcement is witheld [following CRF] subsequent responses are occasionally of longer duration." Notterman (1959) has shown that the mean and variance of the force of bar-pressing in the rat increase from CRF to EXT; Goldberg (1959) has reported comparable findings. Variance ratios for the topography of the vocal response, presented in Col. 4, Table II, show that large increases in response variability were obtained in EXT following CRF in the present study. These findings support both the general statement that response variability increases in EXT after CRF, and also the account of variability among reinforced responses in VI in terms of extinction effects.

The mean parameter values also show an increase in EXT after CRF, with one exception. Prior experimental evidence is scant but tends to support this finding. Margulies (1961) observed an increase in the mean duration of bar-press by the rat in EXT after CRF, as did Hurwitz (1954) and Trotter (1956). The mean force of bar-press also increases in EXT after CRF (Notterman, 1959; Skinner, 1938). Skinner attributes the increase in force that he observed for CRF-EXT animals to "the differentiation of intensity that results from the initial tension of the lever." He goes on to say, "The intensity of the response in an operant is significant only in relation to the differentiative history of the organism." An increase in the mean amplitude of the vocal response was observed in the present study in EXT following CRF and following VI conditioning (Table II, cols. 3, 4). However, the artifact of a manipulandum threshold, which might differentially reinforce response amplitude, was excluded by a suitable adjustment of the sensitivity of the voice-operated relay.

It has been shown that the mean and variance of the parameters of the vocal response increase from CRF to VI-reinforced and from CRF to EXT. There is some evidence for an increase in parameter values within CRF, (Col. 2) although the effect is quite small. Margulies (1961), Notterman (1959), Goldberg (1959), and Antonitis (1950) have obtained the opposite effect: increasing stereotypy during CRF. Table II shows that, in general, the mean and variance of response amplitude, duration and fundamental frequency increase in the order CRF-VI-EXT for Ss 1, 2 and chick, and in the order CRF-CRF-EXT for Ss 3 and 4. Between-groups comparisons are not systematic at all points, but, typically, the mean and variance of the parameter values are greater for reinforced responses in VI after CRF than for the comparable set of responses in CRF after CRF.

The topography of the chick vocal response appears to stand in the same relation to reinforcement operations as that of the human vocal response. The relative mean and variance data for response parameters (Table II) do not provide a basis for discriminating among species of subjects employed.

Summary

The relations among acoustic parameters of the vocal operant are considered and some methods for their measurement described. Four human subjects and one chick are employed in an experiment on the relations among vocal rate, vocal topography and schedules of reinforcement.

The earlier finding that schedules of reinforcement control human and infra-human vocal responding as they do other operants is replicated and extended to the case of variable-interval reinforcement.

An analysis of response amplitude, pitch, and duration shows that the mean and variance of these parameters increases from CRF to VI, from VI to EXT and, for a second group of Ss, from CRF to EXT. The topography of the chick vocal response appears to stand in the same relation to reinforcement operations as that of the human vocal response.

Figure Captions

Fig. 1. Retouched spectrograms of chick (upper) and human (lower) vocal operants during continuous reinforcement (left) and extinction after CRF (right).

Fig. 2. Rates of vocal responding by four human female Ss under three schedules of reinforcement: continuous reinforcement (CRF), variable-interval 64 sec. (VI), and extinction (EXT). The cumulative records for S1 and S2 have been collapsed.

Fig. 3. Vocal responses emitted by one human male S during three successive schedules of reinforcement. The cumulative record has been collapsed.

Fig. 4. Rates of vocal responding by four additional human subjects under three schedules of reinforcement. The cumulative records have been collapsed.

Fig. 5. Rates of vocal responding by one Bantam chick under three schedules of reinforcement. (Only the first 1/5 of the CRF session is shown.)

Footnotes

1. Now at U. S. Army Chemical Center, Edgewood, Maryland.

2. This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education.

3. Also note papers presented at (1) Symposium on the Control of Verbal Behavior, Amer. Assn. for the Adv. of Sci., Denver, 1961, and (2) the Conference on the Experimental Analysis of Behavior, American Psychological Association, Cincinnati, 1959; abstracted in J. exp. Anal. Behav., 1959, 2, 251-269.

4. For a discussion of circuit design, see Peterson and McKinney (1962).

References

Anonitis, J. J. Variability of response in the white rat during conditioning and succeeding extinction and reconditioning. Unpublished doctoral dissertation, Columbia Univ., 1950. Shortened version published J. exp. Psychol., 1951, 42, 217-250.

Delattre, P. The physiological interpretation of sound spectrograms. PMLA, 1951, 66, 864-875.

Fant, G. Acoustic theory of speech production. The Hague: Mouton and Co., 1960.

Ferster, C. B. and Skinner, B. F. Schedules of reinforcement. New York: Appleton-Century-Crofts, 1957.

Flanagan, B., Goldiamond, I. and Azrin, N. Operant stuttering: the control of stuttering behavior through response-contingent consequences. J. exp. Anal. Behav., 1958, 1, 173-177.

Goldberg, I. A. Relations of response variability in conditioning and extinction. Unpublished doctoral dissertation, Columbia Univ., 1959.

Holland, J. G. Technique for behavioral analysis of human observing. Science, 1957, 125, 348-350.

Hurwitz, H. M. B. Response duration of lever pressing in the rat. Quart. J. exp. Psychol., 1954, 6, 62-71.

Ladefoged, P. and McKinney, N. P. Loudness, sound pressure, and subglottal pressure in speech. Unpublished report, Communication Sciences, University of Michigan, 1962.

References (continued)

Lane, H. L. Temporal and intensive properties of human vocal responding under a schedule of reinforcement. J. exp. Anal. Behav., 1960, 3, 183-192.

Lane, H. L. Operant control of vocalizing in the chicken. J. exp. Anal. Behav., 1961, 4, 171-177.

Lane, H. L. Parameters of vowel perception. In: Experimental analysis of the control of speech production and perception. University of Michigan Report 04411-2-P, 1962.

Margulies, S. Response duration in operant level, regular reinforcement, and extinction. J. exp. Anal. Behav., 1961, 4, 317-321.

Millenson, J. R., Hurwitz, H. M. B. and Nixon, W. L. B. Influence of reinforcement on response duration. J. exp. Anal. Behav., 1961, 4, 243-250.

Notterman, J. M. Force emission during bar-pressing. J. exp. Psychol., 1959, 58, 341-348.

Peterson, G. E. and McKinney, N. P. The measurement of speech power. Phonetica (in press).

Schaefer, H.H. & Steinhorst, R.A. The effect of changing the schedule of reinforcement upon duration of responding. J. exp. Anal. Behav., 1959, 2, 335-341.

Shearn, D., Sprague, R. and Rosenzweig, S. A method for the analysis and control of speech rate. J. exp. Anal. Behav., 1961, 197-201.

Skinner, B. F. The behavior of organisms. New York: D. Appleton-Century, 1938.

References (continued)

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

Starkweather, J. A. A speech rate meter for vocal behavior analysis. J. exp. Anal. Behav., 1960, 3, 111-114.

Starkweather, J. A. and Langsley, D. G. Separation of two-channel recordings for vocal behavior studies. J. exp. Anal. Behav., 1961, 4, 193-195.

Trotter, J. R. The physical properties of bar-pressing behavior and the problem of reactive inhibition. Quart. J. exp. Psychol., 1956, 8, 97-106.

Table I

Parameters of response topography for the vocal operant. The mean and variance in amplitude, duration, and fundamental frequency are presented for reinforced responses under CRF and VI schedules of reinforcement and all responses in EXT.

		S_1			S_2			S_3			S_4			Chick		
		CRF	VI	EXT	CRF	VI	EXT	CRF	(CRF)	EXT	CRF	(CRF)	EXT	CRF	VI	EXT
Amplitude (units)	Mean	3.2	4.4	3.8	3.7	5.0	6.3	3.4	4.0	6.1	6.3	6.2	5.5	2.0	2.7	3.8
	Variance	0.4	1.1	1.4	0.7	5.9	7.7	0.3	0.9	0.7	1.0	1.9	3.6	13.0	14.6	60.1
Duration (csec)	Mean	82	84	88	38	51	53	55	55	79	55	54	56	5.9	6.7	11.6
	Variance	459	488	836	87	1259	921	43	48	3398	151	774	2977	21.6	47.4	339.6
Fundamental Frequency (cps)	Mean	264	291	281	224	229	225	216	224	277	267	269	274	2778	2897	3721
	Variance	126	338	470	83	776	353	68	164	1404	124	51	208	109, 018	166, 793	189, 429
Number of Responses		107	1708	1608	81	1950	1906	72	40	248	106	40	46	199	6500	2664
Number of Reinforcements		107	40	0	81	40	0	72	40	0	106	40	0	199	28	0
Time (mins)		15	74	105	15	73	73	15	8	73	15	6	73	15	50	60

	S_5			S_6			S_7			S_8			S_9		
	CRF	VI	EXT	CRF	VI	EXT	CRF	VI	EXT	CRF	VI	EXT	CRF	(CRF)	EXT
Number of Responses	117	3005	8909	98	329	522	134	349	169	111	134	48	95	40	1374
Number of Reinforcements	117	60	0	98	40	0	134	40	0	111	41	0	95	40	0
Time (mins)	16	105	677	15	80	90	15	75	90	15	107	120	4	6	73

Table II

Normalized parameter values for the topography of the vocal operant under schedules of reinforcement. The mean and variance of each parameter measured for each subject during each schedule of reinforcement was divided by the corresponding statistic for that subject obtained during the initial 15 minutes of CRF.

	1			2		3			4	
			Second Schedule					Third Schedule		
		VI		(CRF)				Extinction		
Parameter	S_1	S_2	Chick	S_3	S_4	S_1	S_2	Chick	S_3	S_4
Amplitude										
Rel. mean	1.37	1.32	1.34	1.19	0.99	1.18	1.69	1.88	1.79	0.87
Rel. variance	3.03	9.00	1.13	2.80	1.84	4.00	11.70	4.66	2.40	3.54
Duration										
Rel. mean	1.03	1.33	1.14	1.00	0.99	1.07	1.38	1.98	1.45	1.02
Rel. variance	1.06	14.47	2.20	1.10	5.12	1.82	10.59	15.80	78.66	19.72
Fundamental Frequency										
Rel. mean	1.10	1.02	1.04	1.04	1.01	1.06	1.00	1.34	1.28	1.03
Rel. variance	2.68	9.35	1.53	2.41	0.40	3.70	4.25	1.73	20.65	1.68

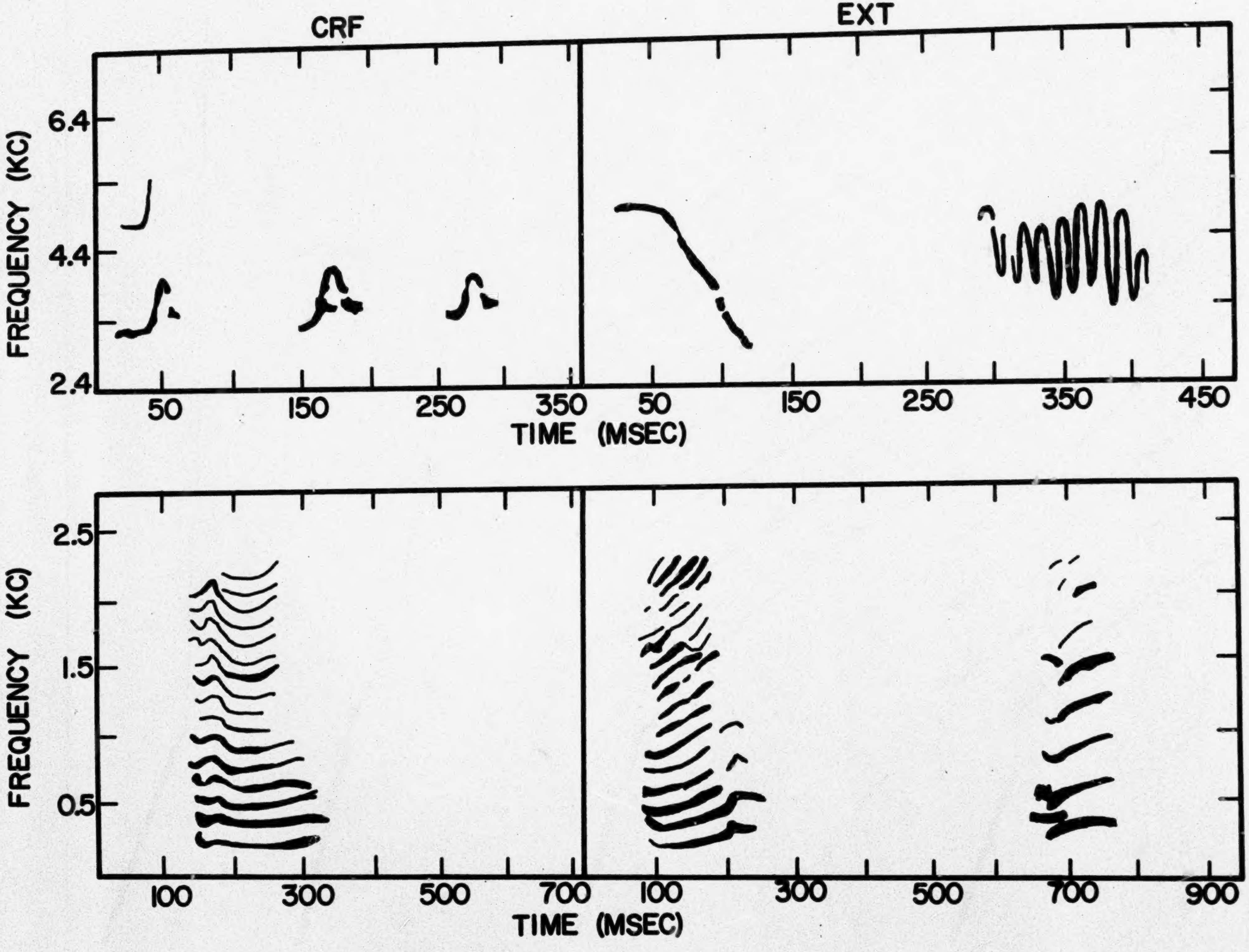
CRF
EXT
FREQUENCY (KC)
6.4
4.4
2.4
50
150
250
350
50
150
250
350
450
TIME (MSEC)
FREQUENCY (KC)
2.5
1.5
0.5
100
300
500
700
100
300
500
700
900
TIME (MSEC)

Fig. 1

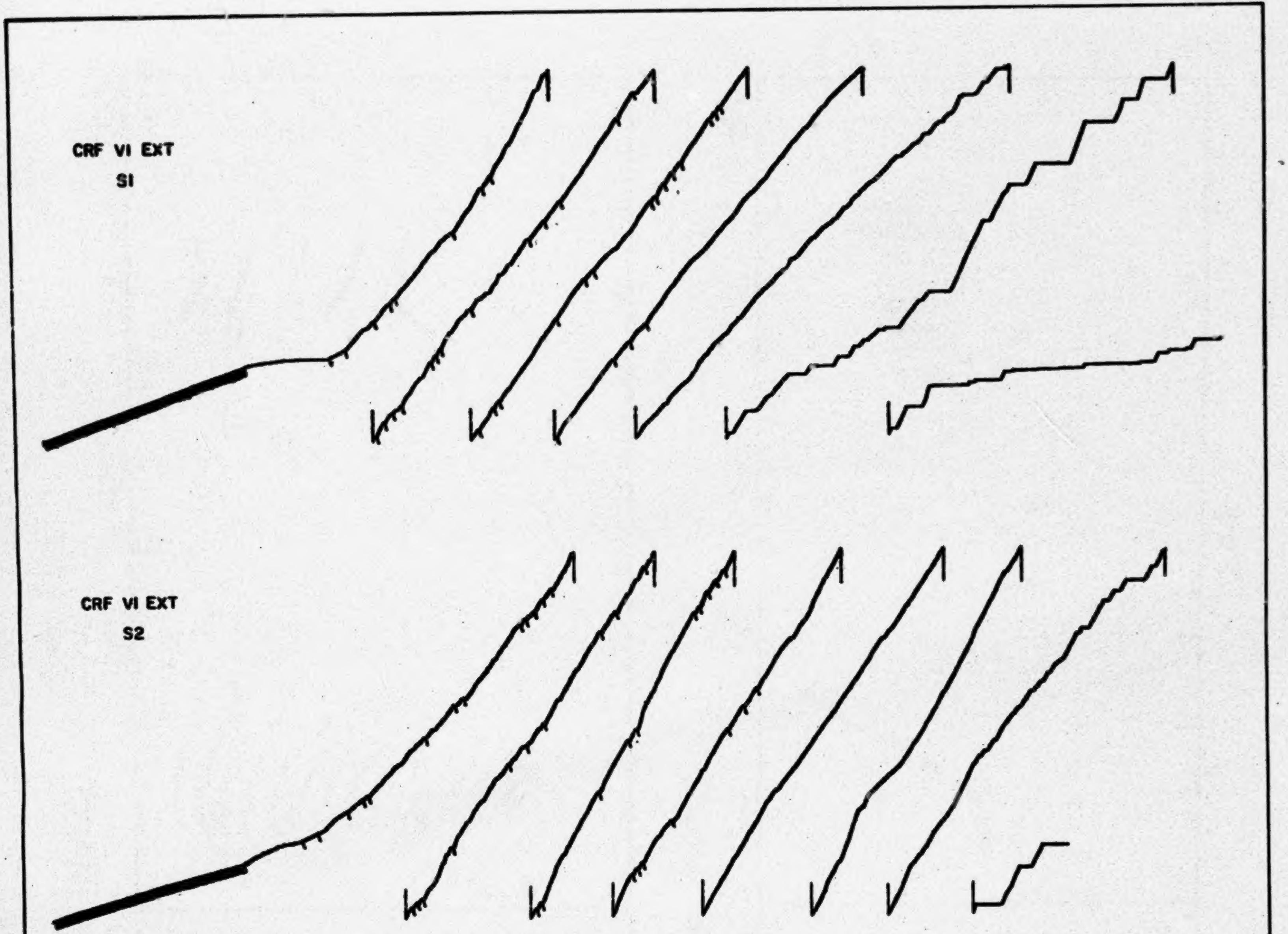
CRF VI EXT
S1
CRF VI EXT
S2
CRF (CRF) EXT
S3
200 RESPONSES
10 MINUTES
CRF (CRF) EXT
S4

Fig. 2

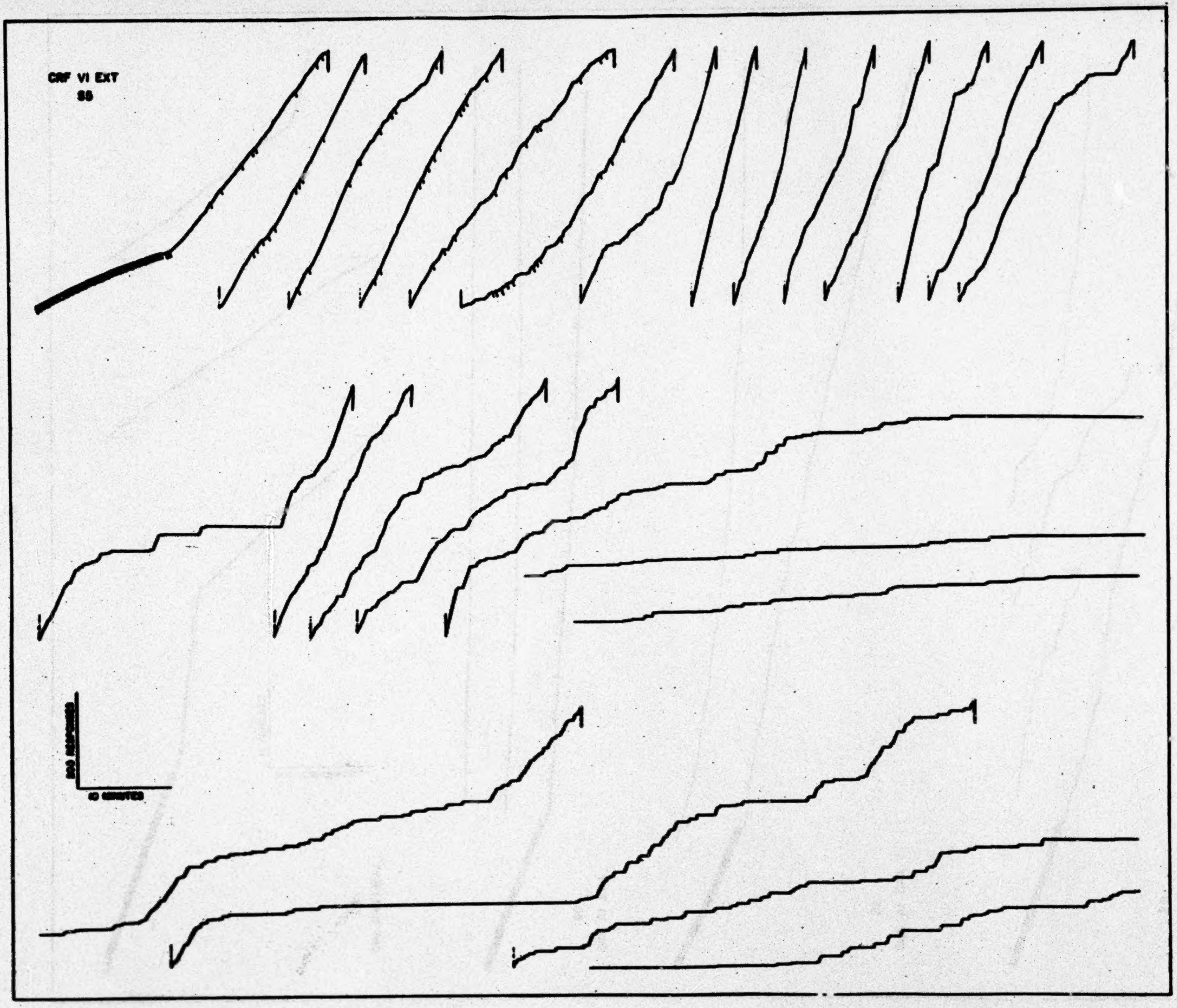

Fig. 3

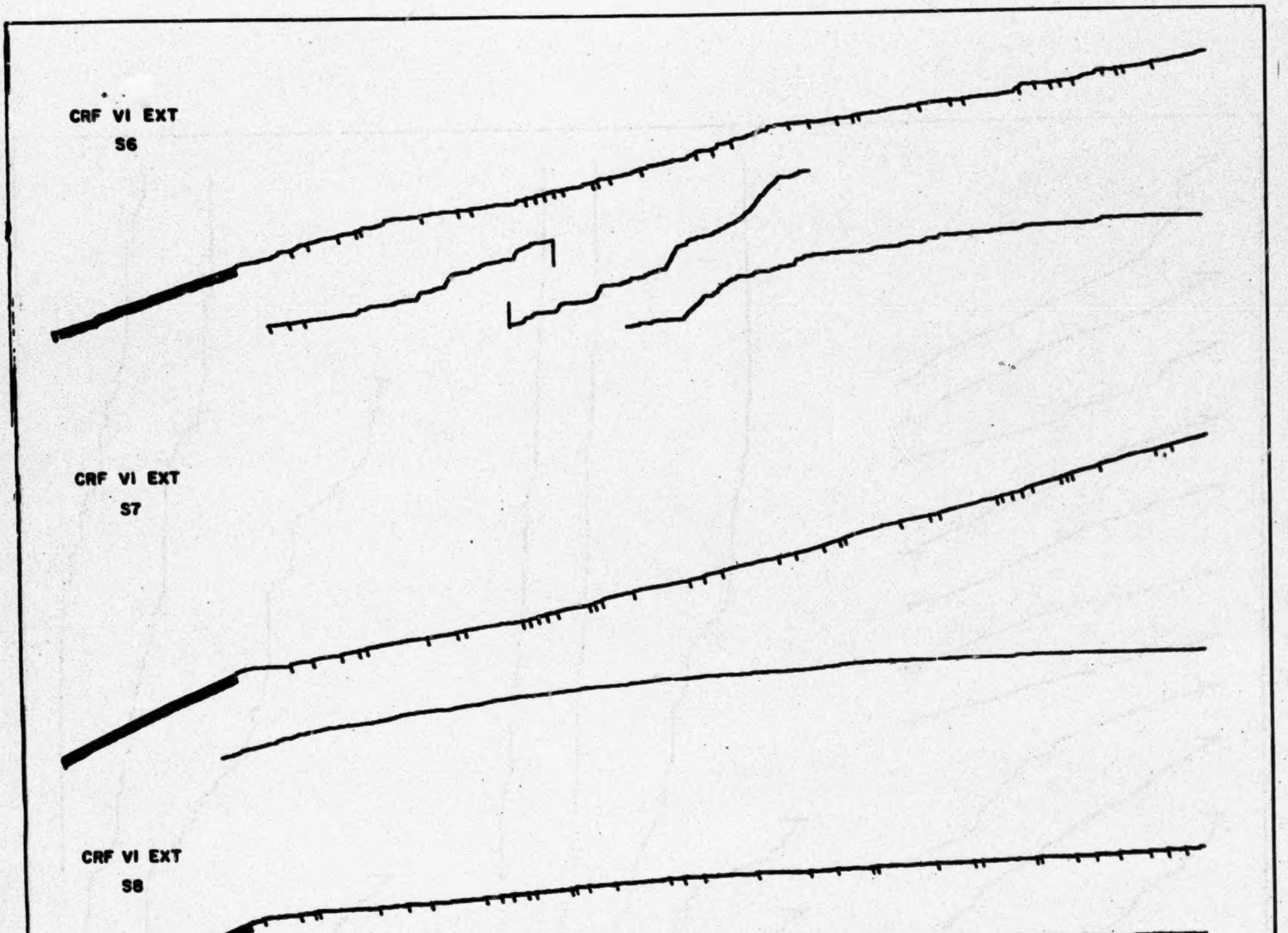
CRF VI EXT
S6
CRF VI EXT
S7
CRF VI EXT
S8
CRF (CRF) EXT
S9
200 RESPONSES
10 MINUTES

Fig. 4

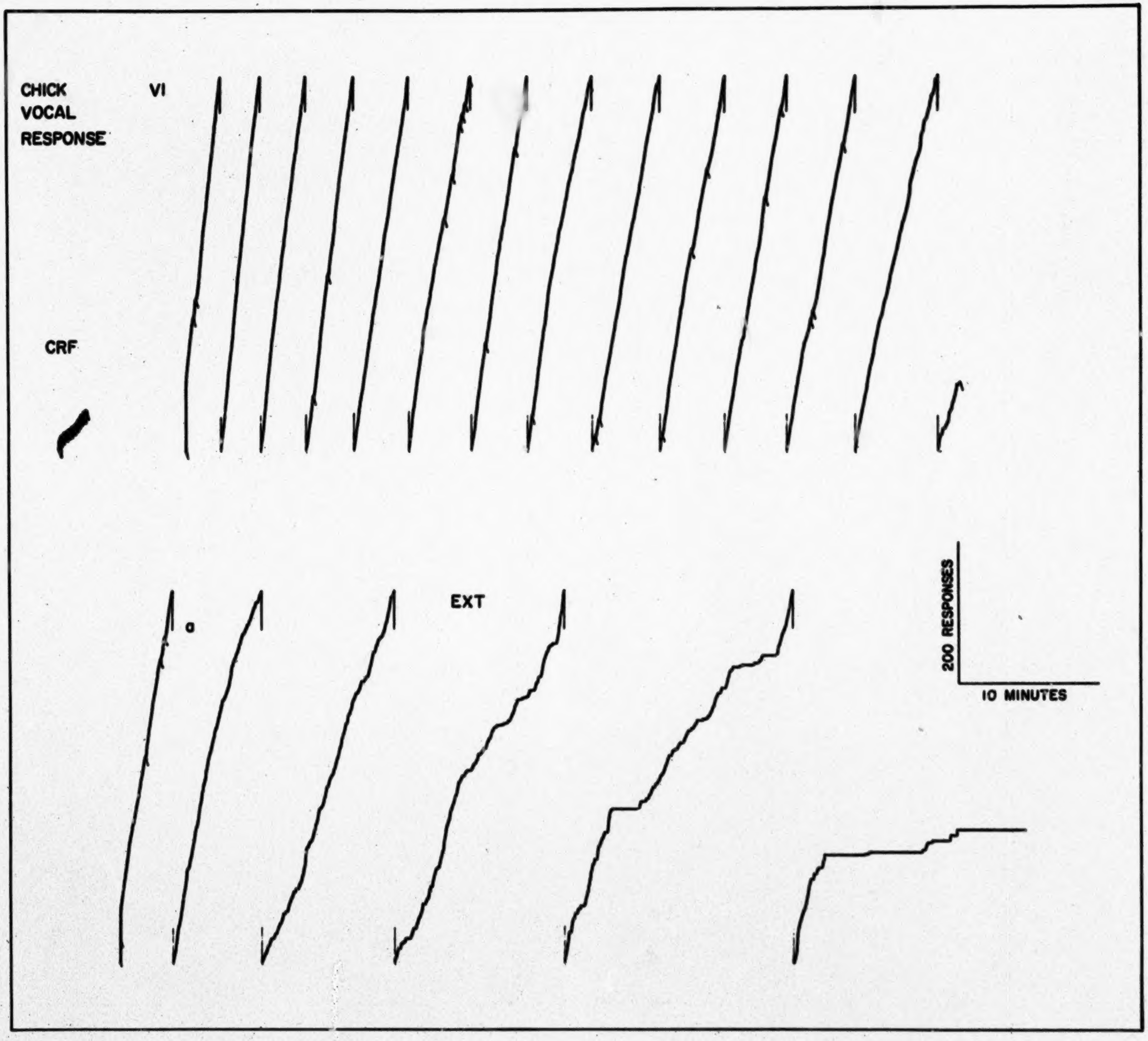

Fig. 5

FOREIGN ACCENT AND SPEECH DISTORTION

Harlan Lane
The University of Michigan

Foreign Accent and Speech Distortion

Harlan Lane
Communication Sciences Laboratory
University of Michigan[1]

The transmission of speech is bounded at both ends by behavior, that of the speaker and the listener. Speech distortion may be broadly defined as any operation that evokes inappropriate behavior by the listener in response to speech. This definition subsumes the wealth of experimental findings obtained when the frequency, amplitude, or time parameters of the speech signal are distorted by the selective action of a transmission system (summarized in Licklider and Miller, 1951). It also includes experimental findings on the influence of context, reinforcement, and physiological variables on the perception of speech.

A metric for the distorting operation is readily provided, when it involves a change in the discriminative stimulus, by specifying the parameters of the transfer function (in cgs units) applied to the original speech signal, or response of the speaker. Other distorting operations may involve a change in concurrent stimulation (Broadbent and Ladefoged, 1960; Skinner, 1936), in the motivation of the listener, in the reinforcement contingencies for his response (Matarazzo, 1961), and in his physiological condition, as in aphasia (Lane and Moore, 1961) or the administration of drugs (Salzinger et al., 1961). Some of these operations are not as readily specified in cgs units, although all may be.

A metric for the effect of the distorting operation, that is, for the behavior of the listener, is provided by 1) selecting the behavioral parameters to be observed, 2) stating the values of those parameters when the listener's behavior is defined as appropriate, 3) employing those values as a reference point for zero distortion, 4) choosing a unit of measurement. In conventional articulation tests (French and Steinberg, 1947), the behavior observed is transcription, appropriate behavior is defined as, roughly,[2] a transcription that matches the pertinent word on the speaker's reading list, and the unit of measurement is the relative frequency of such correct matches. This procedure is based on a nominal scale of speech distortion (Stevens, 1951). An ordinal scale could be obtained from confusion matrices (such as those presented by Miller and Nicely, 1955). Recently, Peterson and Harary (1961) have described a method for measuring phonetic difference that provides a more powerful scale of speech distortion, an interval scale (see later).

Two broad categories of speech distorting operations may be distinguished: response-independent and response-dependent. The former category has received the lion's share of research and includes such operations as filtering, masking, time sampling, etc. These operations are termed response-independent because the parameters of the transfer function applied to a given speech signal are not determined by the probable response of the listener to that signal. Experimental findings show that "vocal communication is highly resistant to distortion" (Licklider and Miller, 1951) of this kind. The situation is otherwise with response-dependent distortion, however. In this category belong those distorting operations that are based on the probable response of the listener during undistorted transmission. The serial transmission of rumor is one example of such distortion (Allport and Postman, 1947). The transmission of an original message by an aphasic or dysarthric speaker is another (Tikofsky et al., 1961). The manipulation of acoustic cues for speech recognition by such devices as PAT (Lawrence, 1953) and Pattern Playback (Cooper et al., 1951) provides a third example. In all these instances, the nature of the distorting operation is most effectively specified in linguistic terms, that is, with reference to the behavior of a standard listener, although, of course, an acoustic transfer function may be written for each speech signal.

The experiments reported in the present article describe some effects of these two kinds of speech distortion and their interaction. Masking and filtering of speech were selected as representative of response-independent distorting operations. Foreign accent was selected as a response-dependent type of speech distortion, and also in view of its practical importance in vocal communication.

Method

Speakers. The four speakers had the following national origins: United States (S_1), Yugoslavia (S_2), India (S_3), Japan (S_4). Thus the Indo-European (Germanic, S_1; Serbian, S_2; Punjabi, S_3) and Japanese (S_4) language groups were represented (Gleason, 1955). Each speaker was male and between 25 and 35 years old. Speaker 1 spoke "General American"; speakers 2-4 had little or no training in spoken English prior to coming to the United States, three months before the experiment. They had a "strong" foreign accent (see listener ratings below) and an inadequate command of English for university study according to the University of Michigan English Proficiency Test.

Word lists. Five "PB" lists of 50 words each were constructed from phonetically balanced sets compiled by the Harvard Psycho-Acoustic Laboratory (Stevens and Beranek, 1942; Egan, 1944). These sets attempt to provide items of monosyllabic structure, equal average difficulty of intelligibility, composition representative of English speech, and words in common usage. The five typewritten lists were presented to each speaker and S_1 read the lists aloud while the group was seated in an audiometric room. Each of the four speakers then read the five PB lists in a different order at the rate of one word every five seconds; 30 seconds elapsed between lists. The spoken lists were recorded on one channel of a four channel tape recorder (Ampex 300-4) and then copied onto a second channel while the record level of each word was adjusted so as to maintain a constant peak amplitude (10 db below 0 VU or approximately 50 db SPL with the TDH-39 earphones employed for listening[3]).

Listeners. Twelve Midwest American undergraduates (six male, six female) served in groups of three in the experiment on speech distortion by masking and foreign accent, and a like number in that on distortion by filtering and foreign accent. The subjects wore binaural calibrated headsets while seated in an audiometric room. Sixty-four stimulus conditions (4 speakers x 4 lists x 4 levels of masking or filtering) were presented in four different hyper-Graeco-Latin square designs, one to each group of three subjects. These designs were orthogonal with respect to masking or filtering. Each listener never heard a speaker read the same list twice, nor was the same list ever heard twice at the same level of masking or filtering. The listeners were told that they would hear common English monosyllables and that they were to were to write them down. The 12 subjects in the first experiment were also presented with a fifth PB list, read by each of the four speakers, and instructed to rate the foreign accent of the speaker on a scale of 1 to 5 ("very little" to "very much"). Finally, this set of four additional PB lists was presented to 222 undergraduates, seated in an auditorium, and to a phonetician for phonetic transcription,[4] in order to compare articulation score and phonetic difference as speech distortion measures.

Masking and filtering. Masking noise was introduced in the first experiment by mixing the tape recorded speech signals with the output of an equal-excitation source (Grason-Stadler, Model 901A) at one of four levels (measured with a Ballantine true r.m.s. VTVM) to give four signal to noise ratios (S/N): 15, 4, -1.5, -5 db. This range of S/N ratios was selected in the light of findings reported by Egan et al. (1943): "When white noise is used...articulation is affected only slightly by S/N ratios greater than +15 db. Further increase in the level of the noise results in a very rapid decrease in articulation until, with a S/N ratio of -10db, articulation is practally zero." The output of the mixer was applied to a low-pass filter (Krohn-Hite 310 AB), with cutoff frequency 8,000 cps (see below) and thence to a low-noise, high-fidelity earphone amplifier that supplied three binaural headsets in parallel.

Frequency selection was employed for response-independent distortion in the second study. The cutoff frequency of the transmission system was set at one of four values: 600, 1200, 2400, or 8,000 cps. To obtain these cutoff frequencies, a sweep-frequency tone was recorded in place of the speech signals, the response of the earphone terminating the system was measured in a 6 cc coupler with a condenser microphone (Western Electric 644A) and graphic level recorder (General Radio, 1521-A), and a suitable adjustment in the nominal filter setting was made. The system had an essentially flat frequency response from the lowest speech fundamental up to the indicated cutoff frequency, whereafter signals were attenuated at the rate of 24 db per octave.

Results and Discussion

Figure 1 shows the effect of signal-to-noise ratio and of foreign accent on the per cent word articulation. The intelligibility of the English speaker decreases at a rate which is comparable to that obtained by Egan et al. (1943) for much larger samples of speakers, words, and listeners. The disparity in the ordinate position of these two articulation curves may be attributed largely to a difference in the received level of speech: 50 and 115 db (SPL), respectively. Egan et al. (1943) have shown that the per cent word articulation is inversely related to the level of received speech at high intensities. Comparable rates of speech distortion as a function of masking were obtained for the foreign speakers, although their articulation scores are, at all points, about 36 per cent below those for the English speaker. It is particularly interesting to note that there is no appreciable interaction effect due to the two types of distortion operating in concert. Furthermore, individual differences among the foreign speakers, expecially with respect to national origin, had no marked effect on articulation scores. The median foreign accent ratings assigned to the four speakers by the twelve listeners were: English, 1; Japanese, 4; Punjabi, 3; Serbian, 3.

Essentially the same relations among response-independent and

response-dependent speech distortion are revealed in Fig. 2, which shows the effect of low-pass filtering and of foreign accent on per cent word articulation. Once again, the articulation curves for the foreign speakers do not differ appreciably from each other but lie, in general, about 36 per cent below that curve for the English speaker. The articulation scores obtained with cutoff frequency 8,000 cps constitute a replication of the first experiment with a second group of 12 listeners; corresponding means do not differ by more than five per cent. As observed in Fig. 1, the two types of speech distortion do not interact in their effects on intelligibility. Articulation scores for a dysarthric speaker are presented for comparison (cf. Tikofsky et al., 1961). Tape recordings of this speaker's rendering of English monosyllables were prepared and presented to the same group of listeners under comparable conditions.

The phonetic difference score associated with each word in the articulation lists rendered by the three foreign speakers was computed according to a method described by Peterson and Harary (1961): "The value [of phonetic difference] is dependent upon the physiological vowel or consonant parameter values by which the two phones differ. In this measure, the vowel and consonant parameter values are roughly scaled according to the magnitude of their separation." The phonetic difference is then specified "in terms of three factors: (a) weighted parameters, (b) the normalized difference between the various parameter values by which the phones differ, and (c) the number of parameter values by which the two phones differ within each parameter."[5]

In terms of the requirements for a distortion metric enumerated

earlier, phonetic difference specifies the topography of the articulatory response as the behavior to be observed. A reference point for the metric is provided by the parameters of articulation inferred from a phonetic transcription of a "standard" speech sample. In the present study, the standard was General American as rendered by S_1. Thus, zero phonetic difference (=zero distortion) would be obtained if the foreign speaker gave the same phonetic rendition of the word as the American. The unit of measurement is based on the parameter values and weightings assigned to phonetic events by the authors.

The mean phonetic difference score for the fifty words rendered by each foreign speaker were: Serbian, 5.1; Punjabi, 3.5; Japanese, 4.5. When the behavior of the foreign speaker is viewed as the terminus of the communication system, these scores are a measure of the distorting effects of his prior verbal conditioning.[6] Liberman et al. (1957) have suggested a way in which this history might operate to distort speech perception: If the listener's speech discriminations "have, by previous training, been sharpened or dulled according to the position of the phoneme boundaries of his native language, if the acoustic continua of the old language are categorized differently by the new one, then the learner might be expected to have difficulty perceiving the sounds of the new language until he has mastered some new discriminations and, perhaps, unlearned some old ones."

When, on the other hand, the behavior of the foreign speaker is interposed in a transmission channel between a native speaker and a native audience, foreign accent may be considered a distorting operation. The phonetic difference scores may then be employed as a measure of the distorting operation, rather than of its effects. In this respect they are comparable to the measures of response-independent distortion employed in the present study: signal to noise ratio and cutoff frequency. Phonetic difference may be related, therefore, to these other measures of speech distortion by determining its effect on some common distortion metric, such as the conventional per cent word articulation. An initial exploration of the relation between phonetic difference and articulation involved computing the phonetic difference scores for each of the fifty words, rendered by each foreign speaker, and then administering an articulation test consisting of tape recordings of these words. The correlation ratio relating phonetic difference to per cent word articulation was -.66 ($p < .01$). Because of the highly skewed distribution of phonetic differences employed and the adverse conditions obtaining during articulation testing, the statistic should be taken to indicate only the feasibility of employing phonetic difference as a measure of response-dependent speech distortion.

Summary and Conclusions

Speech distortion is defined as any operation that evokes inappropriate behavior by the listener in response to speech. Two categories of speech distorting operations are distinguished: response-independent (e.g.,masking, filtering) and response-dependent (e.g., dysarthric speech, foreign accent). Two experiments compare the effects of these two types of distortion on word recognition. Twenty-four Midwest Americans listened to recorded articulation lists rendered by one American and three foreign-born speakers under four conditions of masking or filtering. The phonetic difference between the words rendered in General American and with foreign accent was computed and some properties of phonetic difference as a metric for response-dependent distortion are considered.

1. The intelligibility of the foreign accent speech was approximately 40 per cent less than that of the native speech under all experimental conditions. Differences in intelligibility among the foreign speakers never exceeded ten per cent.

2. A 20 db reduction in speech to noise ratio yields approximately 50 per cent reduction in word articulation for both native and foreign accent speech; the two types of distortion do not interact in their effect on intelligibility.

3. A reduction in the high cutoff frequency of the speech transmission channel from 8,000 to 600 cps yields approximately 40 per cent reduction in per cent word articulation for both native and foreign accent speech; there is no interaction.

4. Phonetic difference may be used as a measure of the effects of speech distortion, or as a metric for the distorting operation in response-dependent speech distortion. An initial test of the relation between phonetic difference as a measure of foreign accent and per cent word articulation gave a correlation ratio of -.66.

Footnotes

1. This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education. The assistance of Mr. K. Anderson, Mr. W. Watrous and Miss A. Crabbs is gratefully acknowledged.

2. The scoring of homonymous forms in transcription as correct reveals that the underlying criterion of appropriate behavior is based on a comparison of the vocal responses of the listener and speaker over an ideal transmission system (cf. the concept of an orthotelephonic system, Inglis, 1938). The use of transcription, therefore, requires several assumptions: "announcers should be selected who are able to enunciate the fundamental speech sounds in a 'normal' manner, where the criterion of normality is one of common sense" (Licklider and Miller, 1951); the listener must employ a standard notation for recording his responses; the experimenter must have some method for establishing equivalences among transcriptions.

3. The level of received speech was measured by impressing an equivalent sine wave voltage (measured on a Ballantine r.m.s. VTVM) on the listener's headphones and measuring the resultant sound pressure level in a 6 cc coupler with a calibrated microphone and VTVM.

4. The assistance of Miss Barbara Erickson is gratefully acknowledged.

5. For parameter values and weightings and a description of the calculational procedure, see Peterson and Harary (1961).

6. An alternate measure, based on word recognition, is the dialect intelligibility ratio, proposed by Lehiste and Peterson (1959).

Figure Captions

Fig. 1. Per cent word articulation as a function of speech-to-noise ratio for native and foreign-born speakers reading English monosyllables. The dotted curve is from an experiment by Egan et al. (1943) in which four natives read 400 to 800 words at each of five S/N ratios to six practiced listeners; the level of received speech was 115 db (SPL). In the present study, speakers of English (triangles), Japanese (open circles), Punjabi (filled circles), and Serbian (squares), each read 200 words at four S/N ratios to 12 listeners; the level of received speech was 50 db (SPL).

Fig. 2. Per cent word articulation as a function of the high cutoff frequency of the transmission channel. Twelve listeners heard one native, one native dysarthric, and three foreign speakers read English monosyllables. Each point is the mean per cent correct transcription for 2400 responses.

References

Allport, G. W. and Postman, L. The psychology of rumor. New York: Henry Holt, 1947.

Broadbent, D. E. and Ladefoged, P. Vowel judgments and adaptation level. Proc. Royal Soc., B, 1960, 151, 384-399.

Cooper, F. S., Liberman, A. M. and Borst, J. The interconversion of audible and visible patterns as a basis for research in the perception of speech. Proc. Nat. Acad. Sci., 1951, 37, 318-325.

Egan, J. P. Articulation testing methods II. Psycho-Acoustic Laboratory, Harvard University, 1944, OSRD No. 3802.

Egan, J. P., Miller, J., Stein, M. I., Thompson, G. G., Waterman, T. H. Studies on the effect of noise on speech communication. Psycho-Acoustic Laboratory, Harvard University, 1943. OSRD No. 2038.

French, N. R. and Steinberg, J. C., Factors governing the intelligibility of speech sounds. J. acoust. soc. Amer., 1947, 19, 90-119.

Gleason, H. A. An introduction to descriptive linguistics. New York: Henry Holt, 1955.

Inglis, A. H. Transmission features of the new telephone sets. Bell syst. tech. J., 1938, 17, 358-380.

Lane, H. L. and Moore, D. J. Operant reconditioning of a consonant discrimination in an aphasic. University of Michigan. Report 04411-1-P.

Lawrence, W. The synthesis of speech from signals which have a low information rate. In Jackson, W. (ed.) Communication theory. London, 1953, 460-471.

Lehiste, I. and Peterson, G. E. Linguistic considerations in the study of speech intelligibility. J. acoust. soc. Amer., 1959, 31, 280-286.

Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C. The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol., 1957, 54, 358-368.

Licklider, J. C. R. and Miller, G. A. The perception of speech. Ch. 26 in Stevens, S. S. (ed.) Handbook of experimental psychology. New York: John Wiley, 1951.

Matarazzo, J. D. Control of interview behavior. Paper read at the Symposium on the Control of Verbal Behavior, Amer. Assn. for the Adv. of Science, Denver, 1961.

Miller, G. A. and Nicely, P. E. An analysis of perceptual confusions among some English consonants. J. acoust. soc. Amer., 27, 1955, 338-352.

Peterson, G. E. and Harary, F. Foundations of phonemic theory. In Jakobson, R. (ed.) Proc. of the Twelfth Symposium on Applied Mathematics. Providence: American Mathematical Soc., 1961.

Salzinger, K., Pisoni, S., Feldman, R. and Bacon, P. The effect of drugs on verbal behavior. Paper read at the Symposium on the Control of Verbal Behavior, Amer. Assn. for the Adv. of Science, Denver, 1961.

Skinner, B. F. The verbal summator and a method for the study of latent speech. J. Psychol., 1936, 2, 71, 107.

Stevens, S. S. Mathematics, measurement and psychophysics. Ch. 1 in Stevens, S. S. (ed.) Handbook of experimental psychology. New York: John Wiley, 1951.

Stevens, S. S. and Beranek, L. L. Word lists for articulation testing. Report No. 1C-26, Psycho-Acoustic Laboratory, Harvard University, 1942.

Tikofsky, R. S., Lehiste, Ilse, and Tikofsky, Rita. A study of the intelligibility of dysarthric speech. Paper read at the 62nd meeting of the Acoust. Soc. Amer., Cincinnati, 1961.

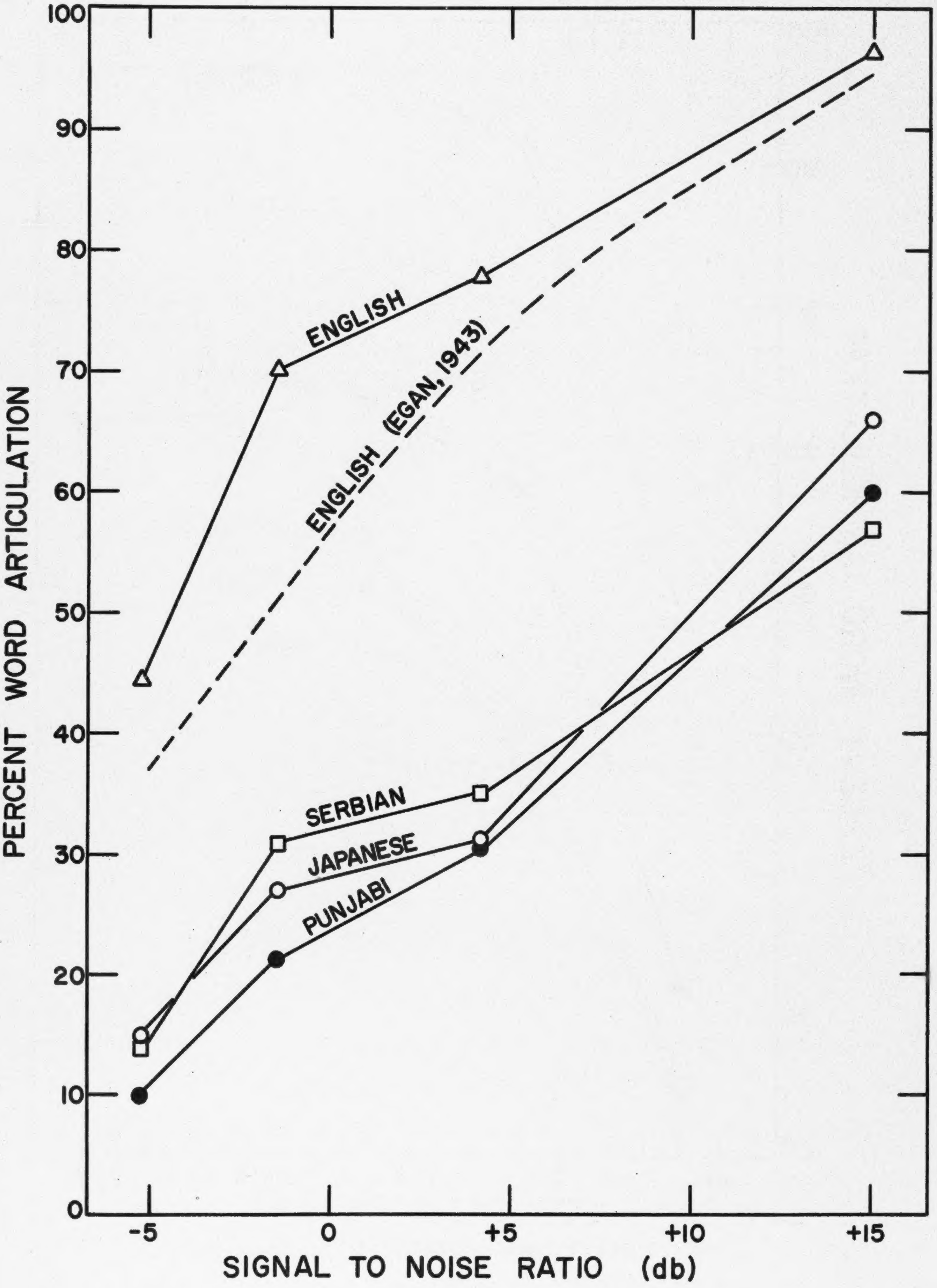
PERCENT WORD ARTICULATION
100
90
80
70
60
50
40
30
20
10
0
-5
0
+5
+10
+15
SIGNAL TO NOISE RATIO (db)
ENGLISH
ENGLISH (EGAN, 1943)
SERBIAN
JAPANESE
PUNJABI

Fig. 1

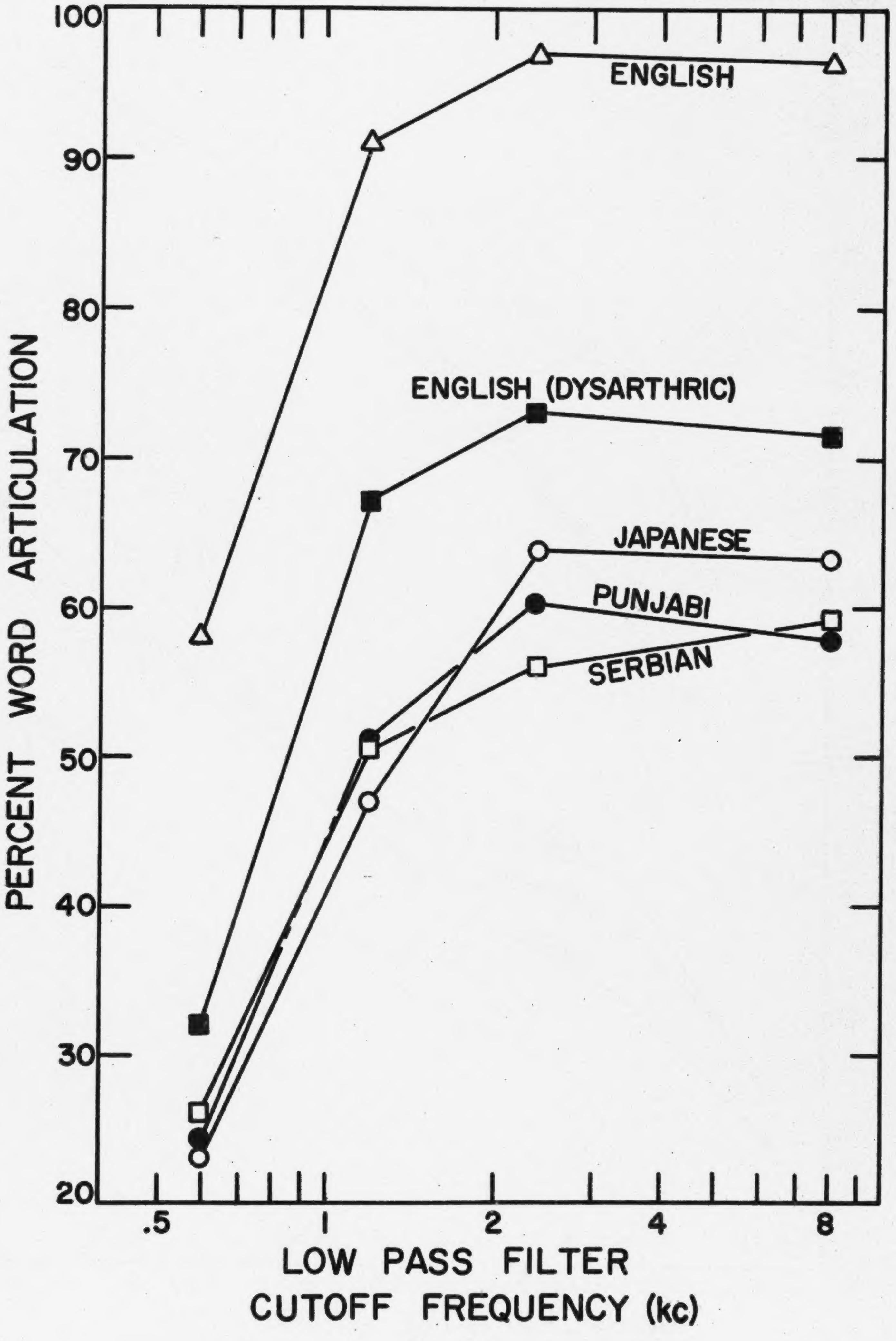
100
90
80
70
60
50
40
30
20
PERCENT WORD ARTICULATION
ENGLISH
ENGLISH (DYSARTHRIC)
JAPANESE
PUNJABI
SERBIAN
.5
1
2
4
8
LOW PASS FILTER
CUTOFF FREQUENCY (kc)

Fig. 2

METHODS FOR SELF-SHAPING ECHOIC BEHAVIOR

Harlan Lane and B. A. Schneider
The University of Michigan

Methods for Self-shaping Echoic Behavior

Harlan Lane and B. A. Schneider
Communication Sciences Laboratory
University of Michigan[1]

An echoic response has been defined as "a response that generates a sound pattern similar to that of the stimulus" (Skinner, 1957). This author observes that an echoic repertoire is established in the child because "it makes possible a short-circuiting of the process of progressive approximation, since it can be used to evoke new units of response..." In the acquisition of a second language, however, this repertoire is only partially effective. On the one hand, it provides echoic behavior that, at the outset, roughly approximates the stimulus. On the other hand, it impedes the development of accurate echoic responding to the extent that the first- and second-language repertories differ. Second-language learning, as a result, must involve both imitation and the reinforcement of progressive approximations to the desired response.

The present study describes six methods for the self-shaping of a minimal echoic operant in a foreign language and the topographical changes in responding that they bring about. The two parameters of echoic responding considered are: pitch slope and duration.

Method

The six conditions of this experiment, incorporating six methods of self-shaping, were: 1) matching only (aural-oral drill); 2) matching-discrimination training-matching; 3) matching-matching with delayed auditory feedback-matching; 4) matching-discrimination training-matching with delayed auditory feedback-matching; 5) matching-matching with visual analog display-matching-free responding; 6) matching-matching with visual digital display-matching-free responding. In each condition the discriminative stimulus was a single Thai toneme /kā/ rendered by a linguist. The relevant parameters of the stimulus were: pitch slope (rate of change in pitch), -13 c/s^2 and response duration, 600 msec.

These matching instructions were read to all subjects:

> "When the experiment begins, you will hear a sound repeated at two-second intervals. Your task is to imitate the sound as accurately as possible between presentations, and to continue to do so until you believe you have given an extremely faithful reproduction. Remember, and I wish to emphasize this point, your task is to reproduce the sound exactly."

In conditions 2 and 4, where discrimination training was employed, a tape recording of five Thai tonemes (/kā/, /ká/, /kà/, /kâ/, and /kǎ/) was presented to the subject. Each toneme appeared eight times in irregular order at four second intervals. Their pitch slopes (c/s^2) and durations (msec) were, respectively, -13, 600; +36, 560; -21, 710; +10 then -120, 650; -39 then +125, 640. The same rendering of /kā/ employed to evoke echoic responses was the positive discriminative stimulus. Discrimination training was continued until S made less than 3 errors in responding to the set of 40 stimuli. These instructions were read:

"During the second phase of this experiment, you will hear a series of stimuli; you are to pull this lever when you hear the first sound in this series and every time afterwards that you hear the same sound. When you do, you will accumulate points on this counter. If you respond to the wrong stimulus, the counter will subtract points. Try to accumulate as many points as you can."

In conditions employing delayed auditory feedback (3,4), the instructions duplicated those for matching. However, S was told that when he responded during one inter-stimulus interval, he would hear his own response played back to him after a slight delay and before the appearance of the next stimulus.

In the condition employing a visual analog display (5), S was seated in front of an oscilloscope (Tektronics 533) that presented two parameters of each response: pitch and duration. These parameters were displayed on the 'scope by filtering the speech signal (80-150 cps; Krohn Hite 310), converting the selected fundamental to a d-c voltage proportional to its frequency (Hewlett-Packard 500 BR frequency meter), and impressing this signal on the vertical axis of the oscilloscope. A one cm. deflection vertically corresponded to a pitch change of four cycles, and one cm. horizontally to a duration of 100 msec.[2] These instructions were read to S:

"In this part of the experiment we would like to give you more information about the sound you have been imitating. It is 600 msec in duration and has a flat pitch slope. You may see these two parameters of your response on the oscilloscope screen in front of you. If your response is like the stimulus, it will trace out a flat line that lasts for exactly six squares. Again, you are asked to continue to imitate the sound between presentations until you are repeatedly able to give a faithful reproduction."

In the condition employing a digital visual display (6), the subject's responses controlled a voice-operated relay (Miratel). A frequency counter (Hewlett-Packard 522B), positioned in front of S, registered the time that the VOR was operated, which equalled the response duration plus 100 msec. The instructions were:

> "In this part of the experiment we would like to give you more information about the sound you have been imitating; it is 700 msec in duration. This counter will help you in giving a faithful reproduction of the stimulus, for it will display the duration of each of your responses in milliseconds. Again you are asked to continue imitating the stimulus until you are repeatedly able to give a faithful reproduction."

In conditions 5 and 6,where "free responding" terminated the session, these instructions were read to the subject:

> "In this part of the experiment you are to continue to repeat the sound until you are giving a faithful reproduction. In this phase, however, you will not hear the stimulus in your headset."

Eighteen subjects served individually under the various experimental conditions in sessions lasting from 20 to 60 minutes. The subject was seated in an audiometric room in front of a microphone and such display equipment as required. He wore a binaural headset with high-fidelity earphones (PDR-8) mounted in doughnut cushions (MX-AR/41), which attenuated air-conducted sidetone by about 15 db.

The discriminative stimulus was first recorded on the fixed diameter loop of a sound spectrograph and then copied repeatedly onto a continuous tape recording (Uher IIIa). This recording was presented in one earphone of the subject's headset while an amount of sidetone was introduced into the other which approximately compensated for the sidetone attenuation due to the headset. The subject's echoic response was recorded on a second tape recorder (Ampex 300) and analyzed, subsequently, in the following manner.

A numerical record of the duration of each response, in milliseconds, was obtained by sending the recorded signals to a calibrated voice-operated relay that controlled an interval timer (Hewlett-Packard 522B frequency counter) and associated printer (Hewlett-Packard 560A). In order to measure the pitch slope, the fundamental frequency was selected from the complex speech signal by filtering (Krohn Hite, 310 AB) and then sent to the frequency counter, which provided a printed record of the period of the voice fundamental at 175 msec intervals, beginning with the onset of the signal. Frequency change, in c/s^2 (pitch slope), was then given by $\frac{\frac{1000}{Tn} - \frac{1000}{To}}{175\ (n-1)}$ where Tn is the terminal period of the voice fundamental, To the initial period and n the number of readings.

Results and Discussion

Figure 1 presents the pitch slope and duration of the echoic responses emitted by each of three subjects in the first four experimental conditions. Figure 2 presents pitch slope and duration for condition 5 (visual analog display), and Fig. 3, for condition 6 (visual digital display). A baseline of echoic accuracy was obtained for each subject in the first phase of each experimental condition. These "matching" data permit the effect of the first-language repertoire to be assessed. The data presented by the unfilled symbols of Fig. 1 and the upper section of Figs. 2 and 3 reveal that, for most subjects, the initial echoic behavior is wide of the mark (broken lines). An estimate of the "linguistically permissible variance," or allophonic variation, may be obtained from the range of pitch slopes and durations given by the Thai speaker in repeated renditions of the discriminative stimulus:[3] duration, 500 to 720 msec; pitch slope, -20 to -5 c/s^2. In simple matching, the response duration of Ss 1, 5, 12, and 15, and the pitch slope of Ss 2 and 15 fall within these boundaries. However, the data for condition 1 show that, in general, this method of "self-shaping" or aural-oral drill as it is more commonly known, does not lead to accurate echoic responding. The duration and pitch slope of the echoic responses tend to stabilize at some value, but this "steady state" does not necessarily have the same acoustic parameters as those of the discriminative stimulus.

The effect of interpolating discrimination training between two periods of echoic responding is shown by the data for Ss 4, 5, 6, (Fig.1). A slight improvement in the correspondence of pitch slope may be noted, but there isno marked effect. Each of the subjects stopped responding sooner in the second matching period than in the first, presumably when he believed he was giving a completely faithful reproduction.

The subject's task of discriminating successive approximations to the desired response is modified somewhat by the introduction of delayed auditory feedback. Under this condition (Ss 7, 8, 9; Fig. 1) S may perform in the manner of a null instrument, modifying his articulation until the paired auditory stimuli (the S^D and his response played back) are no longer discriminably different. There was no evidence, however, of any improvement in echoic accuracy when delayed auditory feedback was introduced.

The methods of discrimination training and delayed auditory feedback were incorporated in condition 4 (Ss 10, 11, 12; Fig. 1). Discrimination training among the tonemes, with their various pitch slopes and durations, might be expected to facilitate the same-different discriminations implicit in the subsequent phase, echoic responding with delayed auditory feedback. In general, the accuracy of echoic responding was improved by these procedures, although not to any marked degree.

The effects on echoic responding of presenting a visual analog of the response parameters are shown in Fig. 2. The upper section of the graph presents the baseline performance of the three subjects. It will be noted that the pitch slope and duration of responding by S_{15} closely approximate the stimulus parameters (broken lines) while those of Ss 13 and 14 do not. The introduction of the display (section 2, Fig. 2) leads to an appreciable improvement in the accuracy of echoic responding for Ss 13 and 14; there was not much room for improvement in the performance of S_{15}. When the display was removed (section 3), echoic accuracy was impaired, although it still exceeded baseline levels for Ss 14, 15. When the stimulus also was removed (section 4), and S instructed to reproduce the sound he had learned, a further reduction in accuracy was observed.

The introduction of the analog display changed the modality sufficient for the discrimination of successive approximations in echoic responding from auditory to visual; it also reduced the large number of dimensions in which the discriminative stimulus could vary, down to the two pertinent dimensions. Although this technique may represent a considerable simplification of the subject's discriminative task, it required, nevertheless, a relatively fine and rapid visual discrimination of a transitory stimulus. In the sixth and final experimental condition, the analog display was replaced in part by a digital display which did not have these limitations. The duration of each of the subject's responses accumulated before him on the frequency counter during voicing and, upon the cessation of voicing, was displayed constantly until the onset of the next response. The effect of the digital display on the correspondence of response duration may be observed in Fig. 3. A marked improvement in echoic accuracy was effected by the introduction of the display (section 2) and maintained following its removal. When both the discriminative stimulus and the display are removed, echoic accuracy is greatly impaired (section 4; cf. Fig. 2). The variance associated with response duration was least during echoic responding with the display and in the post-test immediately thereafter. Variability was greater during simple matching and greatest when both the stimulus and the display were removed.

The efficacy of the digital display may be attributed to the change in the discriminative task afforded the subject following simple matching. In matching without the aid of the display, self-shaping is viable only if S can discriminate successive approximations to the desired acoustic pattern. He must respond selectively, therefore, to the duration of a multi-dimensional stimulus with kinaesthetic as well as air- and bone-conducted sidetone components. The introduction of the display substitutes a visual discrimination among digits, in which S is highly trained, for an auditory discrimination in which he is not. When echoic responding with a high degree of accuracy is desired, this technique may well be extended to other parameters of the echoic response, such as pitch level and relative amplitude.

Summary

Six methods are described for the self-shaping of a minimal echoic operant in a foreign language and their effects on two parameters of stimulus-response correspondence are noted.

1. During self-shaping, in which subjects imitated repeated presentations of a Thai toneme, the duration and pitch slope of echoic vocal responses stabilized at some value. This "steady state" did not necessarily have the same parameter values as those of the discriminative stimulus.

2. Discrimination training, in which the target toneme was contrasted with segments of the same form but different durations and pitch slopes, did not lead to a marked improvement in echoic accuracy.

3. Echoic responding with delayed auditory feedback was not more

accurate than in the absence of this feedback.

4. When the methods of discrimination training and delayed auditory feedback were both introduced, a small improvement in echoic accuracy was noted.

5. Presenting the pitch slope and duration parameters of each response in an analog display led to an improvement in echoic responding that was maintained following the removal of the display.

6. The most effective method for self-shaping of response duration involved the use of a digital display. Echoic accuracy was highest and variability least when the display was present, and directly following its removal. Accuracy was poorest and variability greatest during the pre-test, prior to the introduction of the display, and in a post-test in which both the auditory stimulus and the display were removed. The efficacy of the technique is attributed to the simplification of the discriminative task required in self-shaping.

Footnotes

[1] This research was performed pursuant to a contract with the Language Development Section, U. S. Office of Education. The assistance of Mr. D. M. Brethower is gratefully acknowledged.

[2] Because voicing in /kā/ begins some 60 to 100 msec after the aspirated plosive [k], an echoic response that occupied six cm. on the display was actually slightly longer; a correction was introduced, therefore, in the measurement of response duration.

[3] These were actually interspersed among several renditions of the other tonemic forms of /ka/.

Figure Captions

Fig. 1. Pitch slope and duration of echoic responses by twelve subjects in four experimental conditions. The dotted lines indicate stimulus parameters. The open circles give the mean value of the indicated parameter for blocks of ten responses during simple matching. The filled squares show the values of these parameters obtained under the experimental condition described (see text). Occasionally, the last data point in a set will represent more than 10 but less than 20 responses.

Fig. 2. The effect of a visual analog display of response parameters on the accuracy of echoic behavior. Upper section: mean pitch slope (squares) and duration (circles) of blocks of consecutive echoic responses during simple matching. Section 2: Echoic responding with an analog display. Section 3: echoic responding following the removal of the display. Section 4: reproduction of the vocal response in the absence of the stimulus and display. The dotted lines indicate stimulus parameters. Each point is the mean of n responses, where n is given at the lower right of the graph section.

Fig. 3. The effect of a digital visual display on the duration correspondence of echoic behavior. Upper section: mean duration of blocks of ten consecutive echoic responses during matching. Section 2: echoic responding with a digital display of response duration. Section 3: echoic responding following the removal of the display. Section 4:

Figure Captions
(continued)

reproduction of the vocal response in the absence of the stimulus and display. The dotted line shows the stimulus duration.

Reference

Skinner, B. F. Verbal behavior. New York: Appleton-Century-Crofts, 1957.

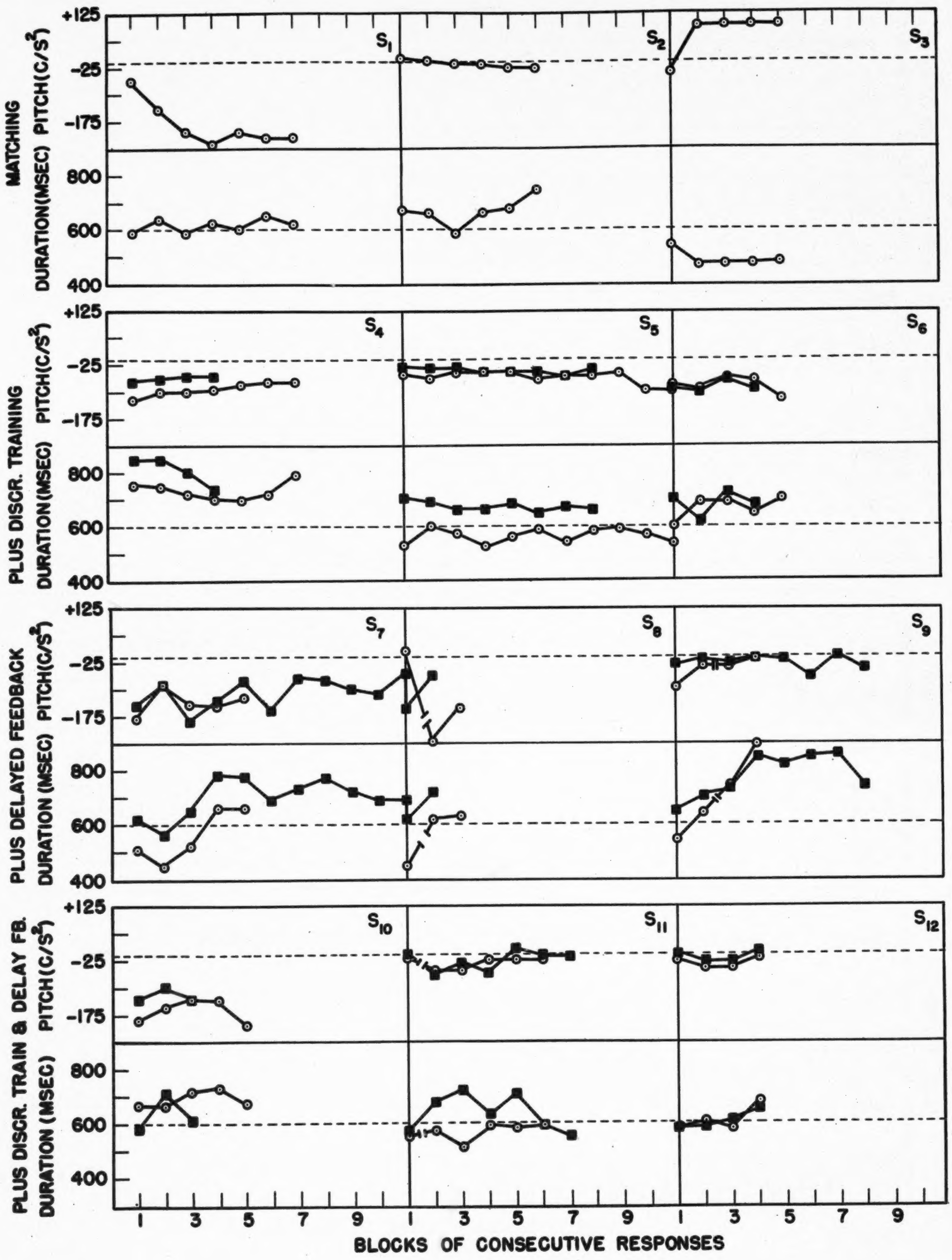
MATCHING
PLUS DISCR. TRAINING
PLUS DELAYED FEEDBACK
PLUS DISCR. TRAIN & DELAY FB.
PITCH (C/S²)
DURATION (MSEC)
+125
-25
-175
800
600
400
S1
S2
S3
S4
S5
S6
S7
S8
S9
S10
S11
S12
1
3
5
7
9
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 1

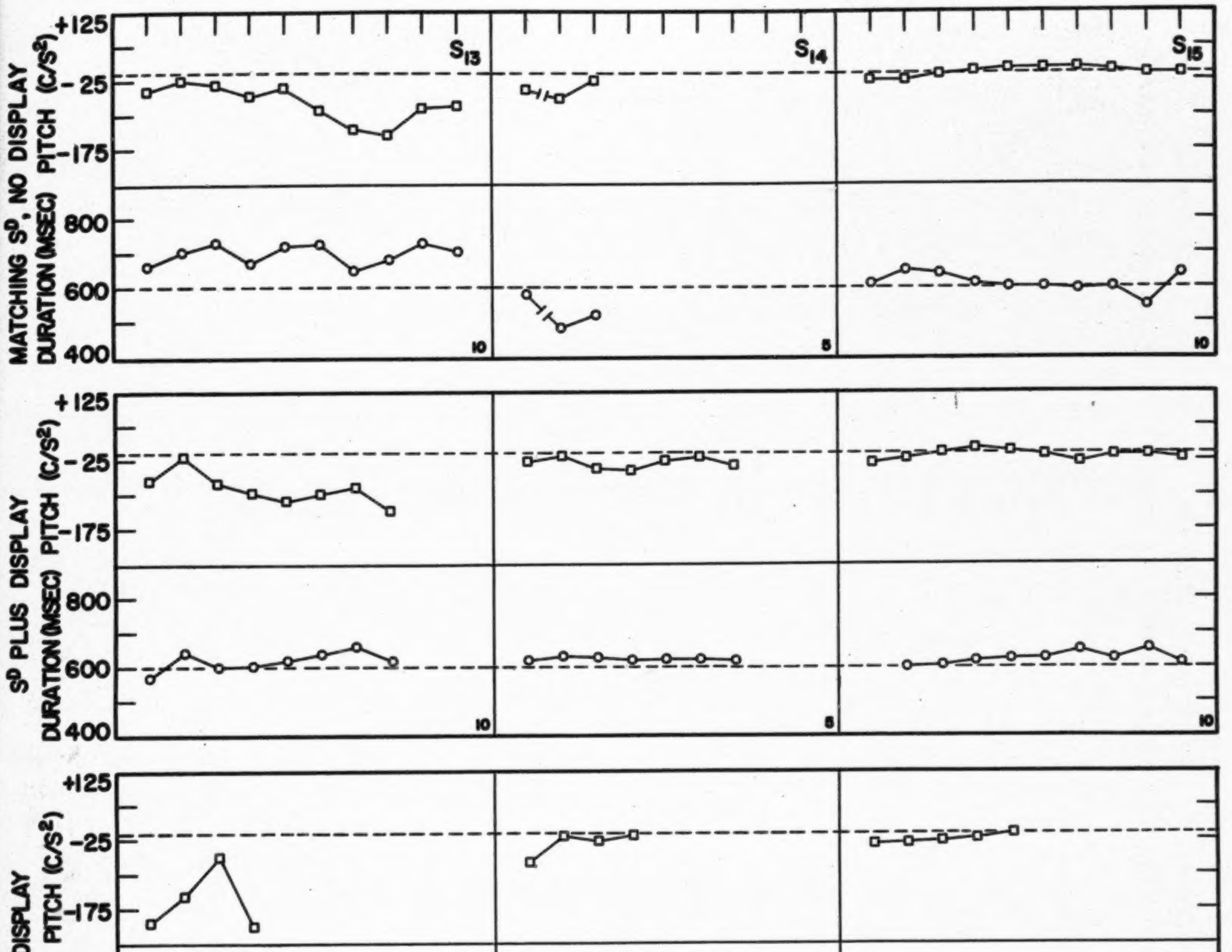
MATCHING S^D, NO DISPLAY
S^D PLUS DISPLAY
S^D, NO DISPLAY
NO S^D, NO DISPLAY
PITCH (C/S²)
DURATION (MSEC)
+125
-25
-175
800
600
400
S13
S14
S15
10
5
>1200
1
3
5
7
9
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 2

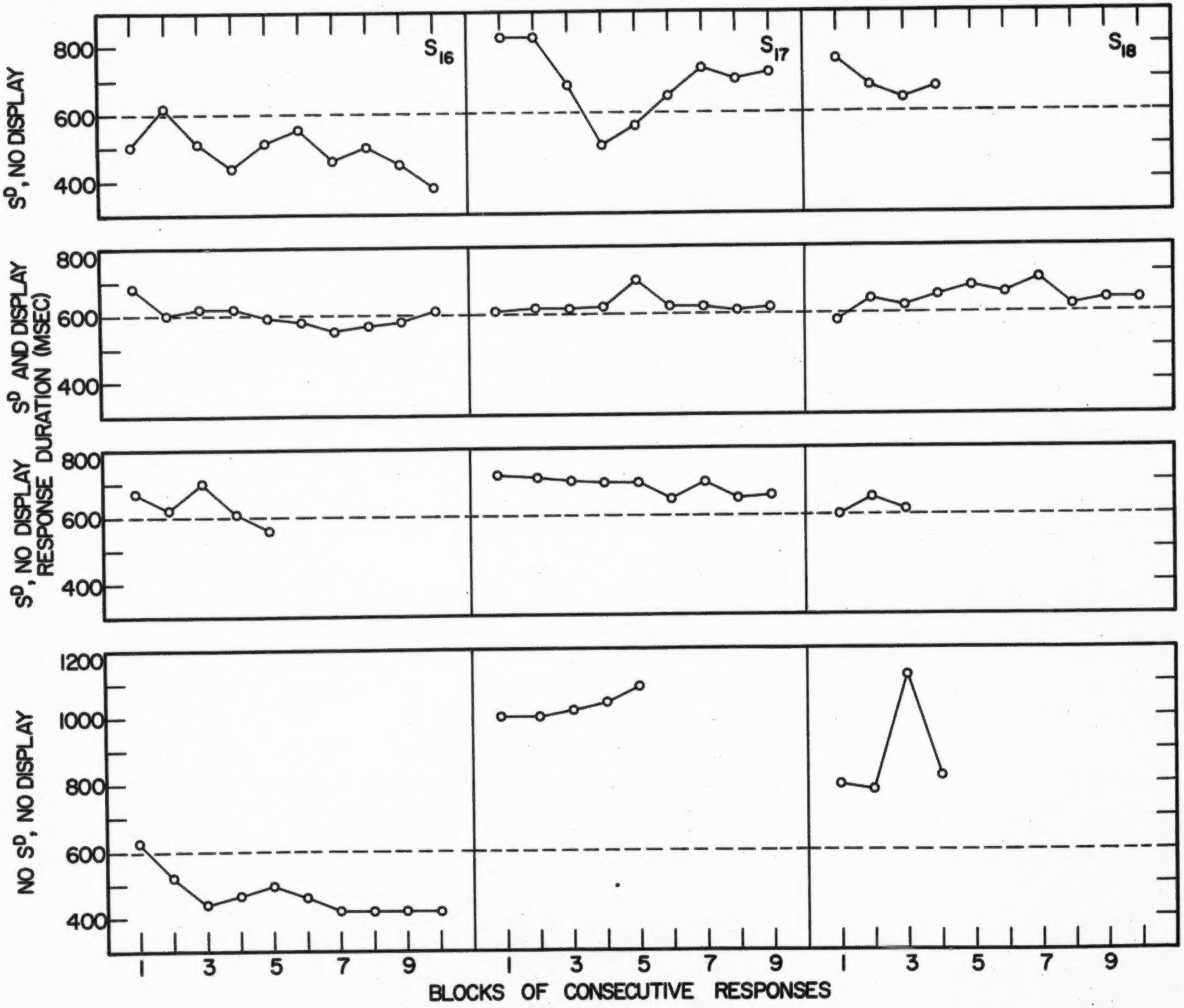
S^D, NO DISPLAY
S^D AND DISPLAY
S^D, NO DISPLAY
NO S^D, NO DISPLAY
RESPONSE DURATION (MSEC)
800
600
400
1200
1000
S_{16}
S_{17}
S_{18}
1
3
5
7
9
BLOCKS OF CONSECUTIVE RESPONSES

Fig. 3

EFFECTS OF TRAINING ON ESTIMATES OF VOWEL LOUDNESS

D. V. Cross and H. L. Lane
The University of Michigan

Effects of Training on Estimates of Vowel Loudness

D. V. Cross and H. L. Lane
Department of Psychology
University of Michigan

In the typical stimulus generalization experiment, we are interested in describing ΔR, the magnitude of decrement of some measured property of the response, as a function of ΔS, an operationally independent measure of the difference between the test stimulus and the initial training stimulus. To make findings based on diverse stimulus continua directly comparable, the stimuli are often described with respect to their spacing on scales of apparent magnitude rather than physical magnitude. These scales are derived from an observer's responses to controlled stimulus variation in psychophysical tasks. Most students of pyschophysics assume, implicitly or explicitly, that there is a linear regression relating R to S', perceived stimulus magnitude, with perfect correlation. There is evidence, however, that this assumed one-to-one correspondence between R and S' does not always hold. Internal conditions and previous experience of the organism, manner of stimulus presentation, and various contextual factors are all determiners of psychophysical judgments (Guilford, 1954). If an experimenter maintains a distinction between the assessment of response effects and the assessment of stimulus, or perceptual, effects in stimulus generalization, then the latter must be independently controlled. In particular, a stimulus generalization gradient may be interpreted or, indeed, represented in terms of the psychophysical scale only if there is some assurance that the particular training procedures involved do not influence the perceived magnitude or discriminability of the stimuli on the continuum employed.

The method of magnitude estimation, a procedure for the direct estimation of psychophysical scales, lends itself to an investigation of the effects of conditioning on the scale of apparent magnitude. This method utilizes a standard stimulus and a set of variable stimuli. The experimenter assigns the numerical value "ten" to a standard of convenient magnitude. The subject assigns a number to the variable that reflects the magnitude of the ratio between variable and standard. Using these methods, Stevens (1958) has demonstrated that, for many sensory continua, the relation between physical and apparent magnitude can be expressed, at least to a first approximation, as a power function of the form $\Psi = S^m$ The value of the exponent m depends significantly on the particular stimulus continuum scaled and is said to represent the operating characteristics of the sensory mechanisms involved.

In investigating the effects of training on magnitude judgments it is important that discriminations along the continuum to be scaled are not reinforced. To do so would perforce alter the psychophysical scale. In the present study, a discrimination task was employed which would permit reinforcement of the modulus response to the stimulus serving as a standard in the subsequent scaling task. This involved using S^{Δ}'s which differed from S^D with respect to qualities orthogonal to the quality to be scaled. We chose a discrimination task involving acoustic stimuli varying in their spectral composition but not in intensity.

Method

The method of magnitude estimation (Stevens, 1958) was used to scale vowel loudness under two experimental conditions: (a) following discrimination training with five vowel sounds, in which the vocal response 'ten' to the middle vowel, /ɛ/, was reinforced, and (b) following a neutral task using the same vowel sounds but without reinforcement of responding. In addition a third condition (c) was employed in which S was simply tested for intensity generalization following the training procedure in condition (a).

The stimuli used in the first part of the experiment consisted of the first three formants of the vowels: /i/, /I/, /ɛ/, /æ/, and /ɑ/. These were electronically synthesized using the formant frequencies and relative formant amplitudes presented by Peterson and Barney (1952). A calibrated oscillator (Hewlett-Packard 200) drove a pulse generator that supplied the fundamental frequency and its harmonics to a narrow band-pass filter (Dytronics) which was tuned, for each vowel, to pass in turn the first, second, and third formant frequencies for recording on separate channels of a four-channel tape recorder (Ampex 300-4). These formant signals were then separately amplified on playback and mixed to obtain the desired vowel sounds. A sequence of 100 stimuli, each with one second duration and 50 msec rise and decay time, were then recorded on magnetic tape for presentation at six-second intervals. Each of the five vowel sounds occured an equal number of times in a randomized order. The recording levels were adjusted so that on playback all stimuli would have the same VU level.

Thirty-three naive undergraduates served individually in sessions lasting 22 minutes. The twelve Ss in group I were given a sheet on which was printed 100 lists containing the words: heed, hid, head, had and hod. S was told that this was an experiment in speech perception and that he would be presented a randomized sequence of synthesized vowel sounds. He was instructed simply to check, on the sheet provided, the word containing the vowel sound he heard upon each presentation. The 11 Ss in group II and 10 Ss in group III were read the following instructions:

> "This is an experiment in the perception of speech. You will be presented a random sequence of synthesized vowel sounds. We want you to learn to identify one of them by saying 'ten' each time you hear it. If you are right, ten points will be automatically added to the score displayed in front of you. A response to any sound other

> than the correct one, as well as failure to respond when it is presented, will be an error and will not add to your score. We want to find out two things: (1) how long it will take you to learn this identification and (2) how many errors you will make. So try to get a high score, making as few errors as possible."

The stimuli were presented to S monaurally through a calibrated headphone (PDR-8) at a sound-pressure level of 75 db (re. 0.0002 dyne/cm^2). An Ampex 300-4 tape recorder was used with an electronic switch and timer (Grason-Stadler Model No. 829 S119) which was triggered by synchronizing pulses from an additional track of the stimulus tape. The switch was used to pass only the recorded stimuli, eliminating tape noise and print-through occurring between stimuli.

Upon completion of this phase of the experiment, Ss in groups I and II were given the following instructions for making loudness judgments:

> "The vowel /ɛ/ will be presented at various intensity levels. Your task is to estimate its loudness at each of these levels by assigning it a number. If it seems to have the same loudness as in the previous task, which we will refer to as the 'standard' loudness, call it '10'. If it seems 'louder' or'softer' try to assign a number to it which represents 'how much' louder or softer it is relative to the standard, that is, a number proportional to 10. For example, if it seems twice as loud call it 20, if half as loud call it 5, etc. You may use any number that seems appropriate, whole numbers, fractions, or decimals. The standard will be presented five times before the regular series begins."

In addition, Group II Ss were told that, although their own counter would be inoperative, their score would continue to accumulate on E's counter; i.e., if they correctly identified the modulus stimulus by calling it 'ten' they would still get their points. The Ss in group III were told that the vowel /ɛ/

would be presented at various intensity levels and they were to respond as before if it seemed to be the same level as that during training; otherwise they were not to respond. They were also told that they could continue to earn points for correct responses, but only on E's counter. These Ss were presented the test series directly without the five preliminary presentations of the standard intensity level.

The test stimuli were 1-sec presentations of the synthesized vowel /ɛ/. They were recorded on magnetic tape at 5 db intervals over a 50 db range. The modulus stimulus was set at 75 db, the same level used in the preceding task. Ten quasi-random permutations of the eleven stimuli were presented in order for a total of 110 stimulus presentations. The constraints placed upon their order was that each stimulus followed every other stimulus exactly once. This was done in order to counterbalance anchoring effects.

Results

Figure 1 shows the mean log estimates of loudness plotted separately for groups I and II. The straight lines drawn through the points represent a least squares fit for the function $\log R = \log a + b \log S$. It is apparent that a power law describes the obtained data well. There are no systematic departures from linearity except for the slight curvature at the top of the group II function. The slope for group I is 0.47 and for group II, 0.62. A Mann-Whitney U-test, computed on the slope parameters determined separately for each individual, yielded $U = 29.5$. For samples of this size, a U this large or larger would be expected by chance less than 2.5% of the time.

The results for individual Ss are presented in Fig. 2. The points in these plots represent the mean log estimates as determined from 10 judgments at each intensity level for each subject. The straight line drawn through each individual's plot, determined by a least squares fit to the data, represents the linear component of his loudness function. It is apparent that inter-individual variability in the slopes of these functions is high, although the between-groups difference is clear. There was virtually no overlap between the interquartile ranges for the two groups. The median and lower and upper quartile exponents for group I are: mdn = .42, Q_1 = .37 and Q_3 = .51; and for group II: mdn = .66, Q_1 = .51 and Q_3 = .72. For all plots, the linear component of trend accounts for the maximum amount of variance.

Assessing the overall effects of previously reinforcing the modulus response, R^M, it was observed that R^M was emitted with an average relative frequency by each S for groups I, II, and II respectively of 13.0, 16.0, and 20.2. In Fig. 3 are presented generalization gradients for all three groups broken down into successive fifths of the testing sessions The ordinate values represent the average frequency of the modulus response (ten) to each stimulus intensity. Comparing groups I and II it is apparent that prior conditioning resulted in greater consistency for the gradient to peak at S^m. Group I emitted approximately the same average number of modulus responses during the last part of the session as the first, but variability, and thus inaccuracy, increased considerably. Variability also increased for Group II but the number of responses emitted increased proportionally, thus relative variability was less. A similar finding holds for group III; however, these gradients peaked consistently at a stimulus intensity of 80 db rather than the initial training intensity. In interpreting this result it must be remembered that group III was not presented the standard intensity again prior to testing, as were groups I and II.

Discussion

It has been demonstrated clearly that prior conditioning alters the scale of subjective magnitude, as constructed from direct judgments of vowel loudness.

This is tantamount to saying that, if S is reinforced for emitting the modulus response in the presence of the standard stimulus, he will assign larger numerals to intense stimuli and smaller numerals to weak stimuli than he would if he had not been reinforced. Which of the two functions obtained, it may be asked, best describes the "true" scale of vowel loudness? The function obtained from group I has an exponent (slope) of 0.47 which is close to 0.4, the exponent reported by Lane (1961) for the loudness of the synthesized vowel /a/. On the other hand, the function obtained from group II has the exponent 0.62 which is close to that reported by Stevens (1955) for the loudness of a 1000 cps tone. Stevens investigated the method of magnitude estimation thoroughly to identify sources of potential bias and pitfalls in its use (Stevens, 1956). Were any of thse biases operating here which could explain the differences obtained between groups? Although the scaling procedure employed may not be generally recommended, both groups did receive comparable experimental treatment in magnitude estimation: the same standard intensity level, same range and stimulus order, and the same instructions. The prior training given group II did not involve intensity discriminations, as all stimuli were equated in this respect. Further, it is difficult to see how reinforcing the vocal response "ten" could operate on the *a priori* probabilities of emission for other numerals in the response repertory.

Similar problems of interpretation are raised when individual differences are considered. The individual loudness functions obtained in this study are orderly, implying a consistency in individual judgments--but the empirical constants obtained by repeated measurement of the same S are not the same as those obtained by averaging over Ss. That is, each S seems to have his own personal loudness scale if a strict interpretation is imposed upon this scaling procedure. Jones and Marcus (1961) found significant subject effects in magnitude judgments in three modalities. This implies a consistency in an individual's use of numbers that holds across modalities. Either Ss differ in their concept of the number system, i.e., in their habitual ways of using numbers, or their sensory tranducers differ, or both. The first alternative implies that an individual's judgments do not, in fact, reflect "real" metric properties of sensory magnitudes. The second alternative implies that the sensory mechanisms involved have different operating characteristics for different individuals. If we accept the highly deterministic rationale underlying the direct scaling procedures we are forced to accept this onerous second alternative. The dilemma may be resolved by considering a third alternative suggested by the following argument.

When S judges the apparent magnitude of a variable stimulus with respect to a given standard he is, in fact, basing his judgment on a remembered standard. Seldom, if ever, are two different intensities presented simultaneously. In the present study, where the standard intensity was presented only at the beginning of a relatively long stimulus series, S's recollection of the standard was particularly taxed.

Memory of stimulus intensity may be presumed to "drift" in time and is subject to systematic biases that have been studied in connection with the time-order error in psychophysics, context and anchoring effects, adaptation level, and judgmental relativity (Stevens, 1958a). Comparison of the group I and group II generalization gradients shows that the discrimination training given group II Ss in the present study resulted in more orderly generalization gradients that peaked sharply at the modulus stimulus. Inference from this finding suggests that the variability in the remembered modulus was less for group II Ss.

Variability in the "remembered" standard and anchoring effects may operate in concert to produce slope differences in scaled judgments. If a variable stimulus effectively anchors the "remembered" standard, causing it to drift in the direction of the variable, (so that the standard is remembered as being "more like" the variable than it actually is) then the judged magnitude of the variable will be greater, or less, depending upon whether the variable is less or more intense than the standard. In other words, if a variable is actually five times greater in subjective magnitude than the true standard, but it is compared to a remembered standard which has drifted in a direction toward the variable, for example, to a level for which the variable is only four times greater, the judged magnitude would reflect this and, in effect, underestimate the predicted scale value. This interpretation of the difference between the group I and II loudness scales would account for the results obtained here and also for individual differences in scaling since it is reasonable to assume that individuals differ with respect to memory characteristics and susceptibility to external anchoring.

Summary

The effects of prior conditioning on judgments of subjective magnitude were assessed using the method of magnitude estimation to scale vowel loudness under two experimental conditions: (a) following discrimination training utilizing five synthesized vowel sounds (/i/, /I/, /ɛ/, /æ/ and /ɑ/) for which the vocal response "ten" to the middle vowel /ɛ/ was reinforced, and (b) following a neutral task using the same vowel sounds but without reinforcing differential responding. In addition, generalization of the modulus response along the intensity continuum was compared to that resulting when magnitude judgments are not required. The loudness scales obtained under the first two conditions confirmed the power law. Prior conditioning resulted in a function with a significantly steeper slope than that obtained under neutral conditons. This result was discussed in terms of a memory mechanism and anchoring effects and extended as an explanation of individual differences in judgmental scales.

Figure Captions

Fig. 1. Effects of prior training on the vowel loudness function. Each point represents a mean log estimate based on ten judgments from each of 12 Ss in group I and 11 Ss in group II. The straight lines drawn through the respective plots are described by $\Psi = kS^{0.62}$ (filled circles) and $\Psi = kS^{0.47}$ (open circles).

Fig. 2. Vowel loudness functions for individual Ss. Each point is the mean log estimate of loudness based on ten numerical estimates by each S. Group II Ss (N=11) received prior reinforcement of the modulus response; group I Ss (N=12) received a prior neutral task.

Fig. 3. Auditory generalization gradients for the modulus response partitioned by experimental conditions and by successive blocks of stimulus presentations. Each point represents the average frequency of the response "ten" on two presentations of the indicated stimulus for the Ss of group I (N=12), group II (N=11) and group III (N=10).

References

Guilford, J. P. *Psychometric Methods*, New York: McGraw-Hill, 1954.

Lane, H. L. The effects of changing vowel parameters on perceived loudness and stress. In: *Experimental analysis of the control of speech production and perception*. Ann Arbor: University of Michigan Res. report 04411-1-P, 1961.

Peterson, G. E., and Barney, H. L. Control methods used in a study of the vowels. *J. Acoust. Soc. Amer.*, 1952, 24, 175-184.

Stevens, S. S. The measurement of loudness. *J. Acoust. Soc. Amer.*, 1955, 27, 815-829.

Stevens, S. S. The direct estimation of sensory magnitudes. *Amer. J. Psychol.*, 1956, 69, 1-25.

Stevens, S. S. Problems and methods of psychophysics. *Psychol. Bull.*, 1958, 54, 177-196.

Stevens, S. S. Adaptation-level vs. the relativity of judgment. *Amer. J. Psychol.*, 1958a, 71, 633-646.

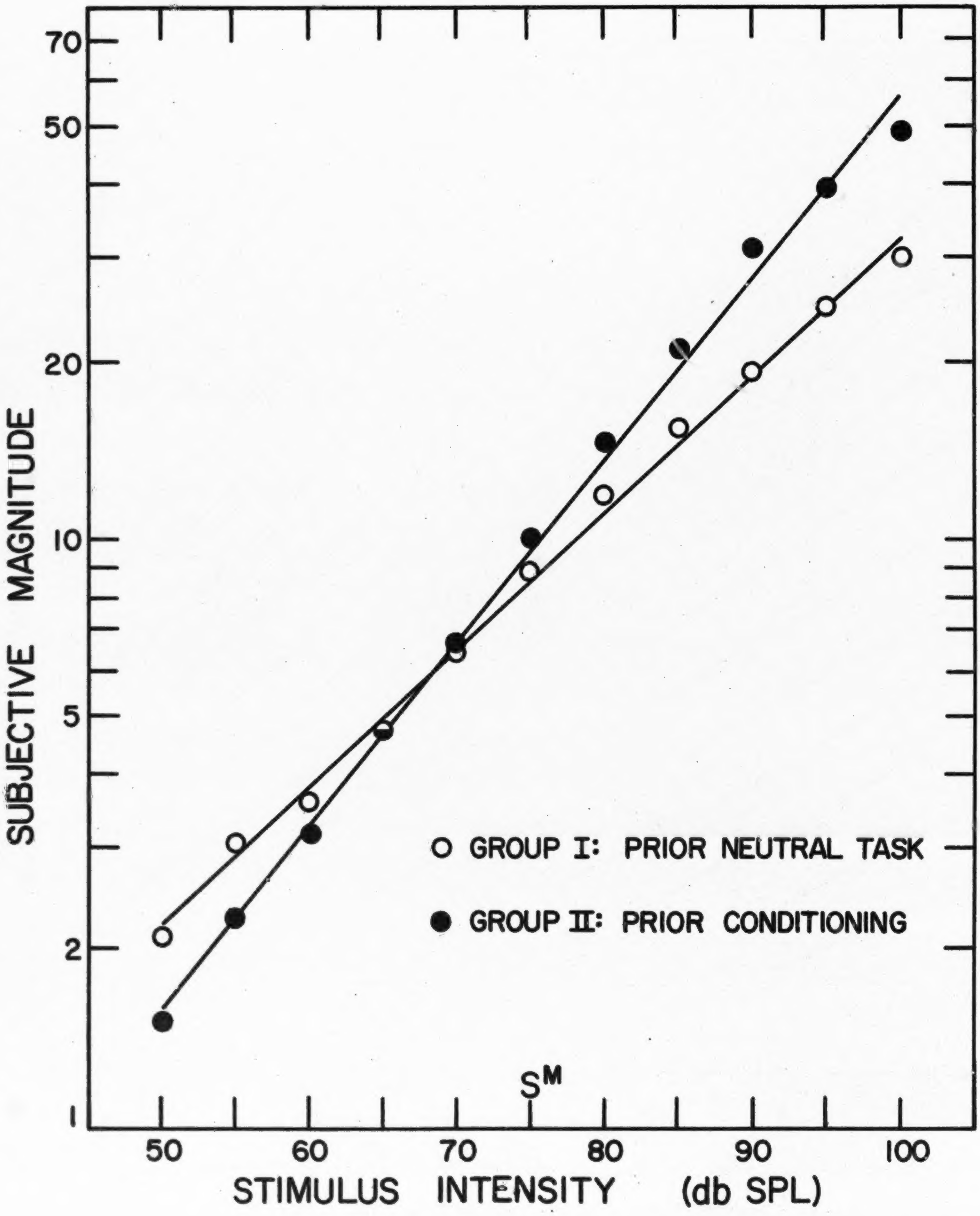
70
50
20
10
5
2
1
SUBJECTIVE MAGNITUDE
○ GROUP I: PRIOR NEUTRAL TASK
● GROUP II: PRIOR CONDITIONING
S^M
50
60
70
80
90
100
STIMULUS INTENSITY (db SPL)

Fig. 1

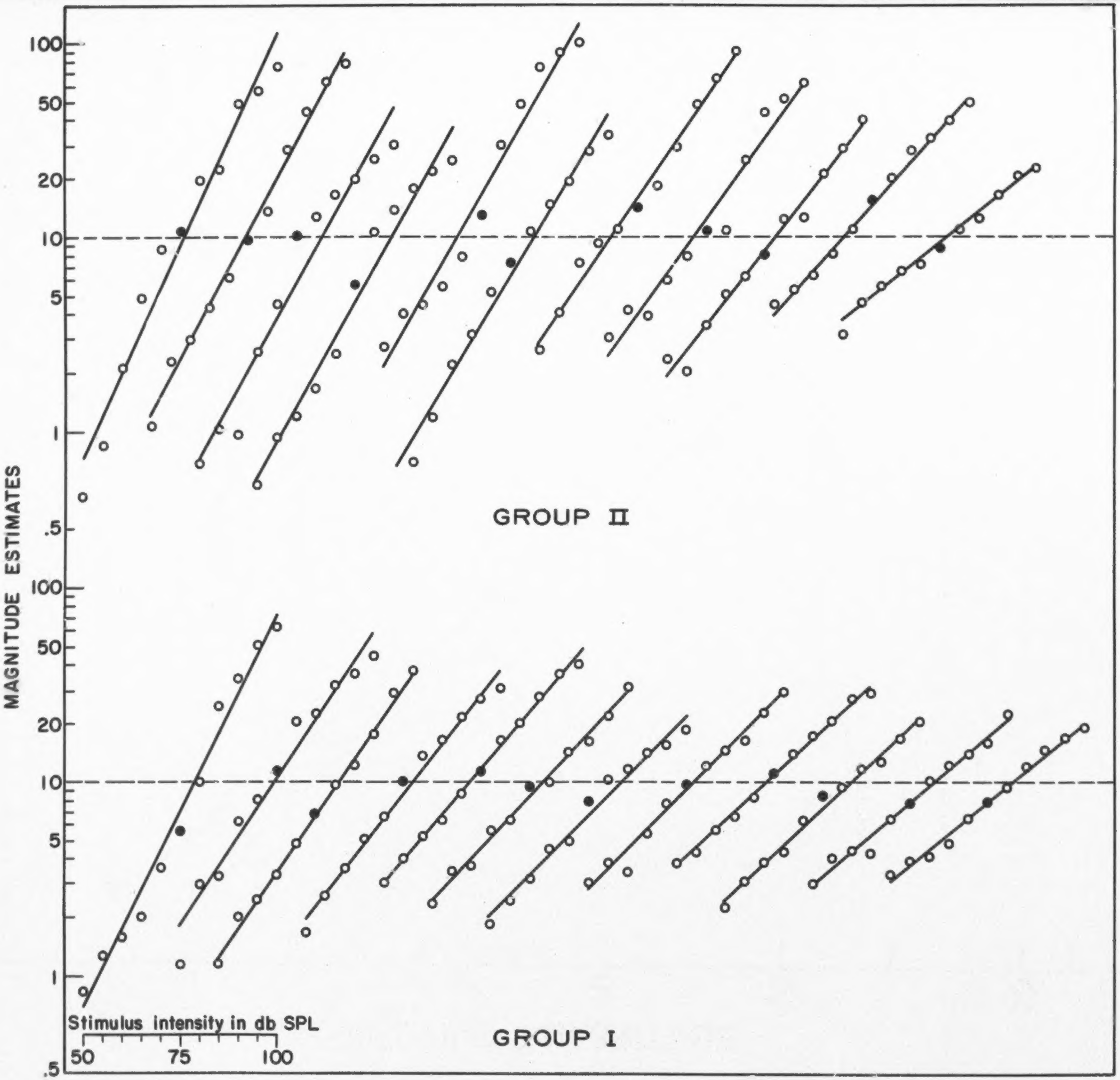
100
50
20
10
5
1
.5
100
50
20
10
5
1
.5
MAGNITUDE ESTIMATES
GROUP II
GROUP I
Stimulus intensity in db SPL
50
75
100

Fig. 2

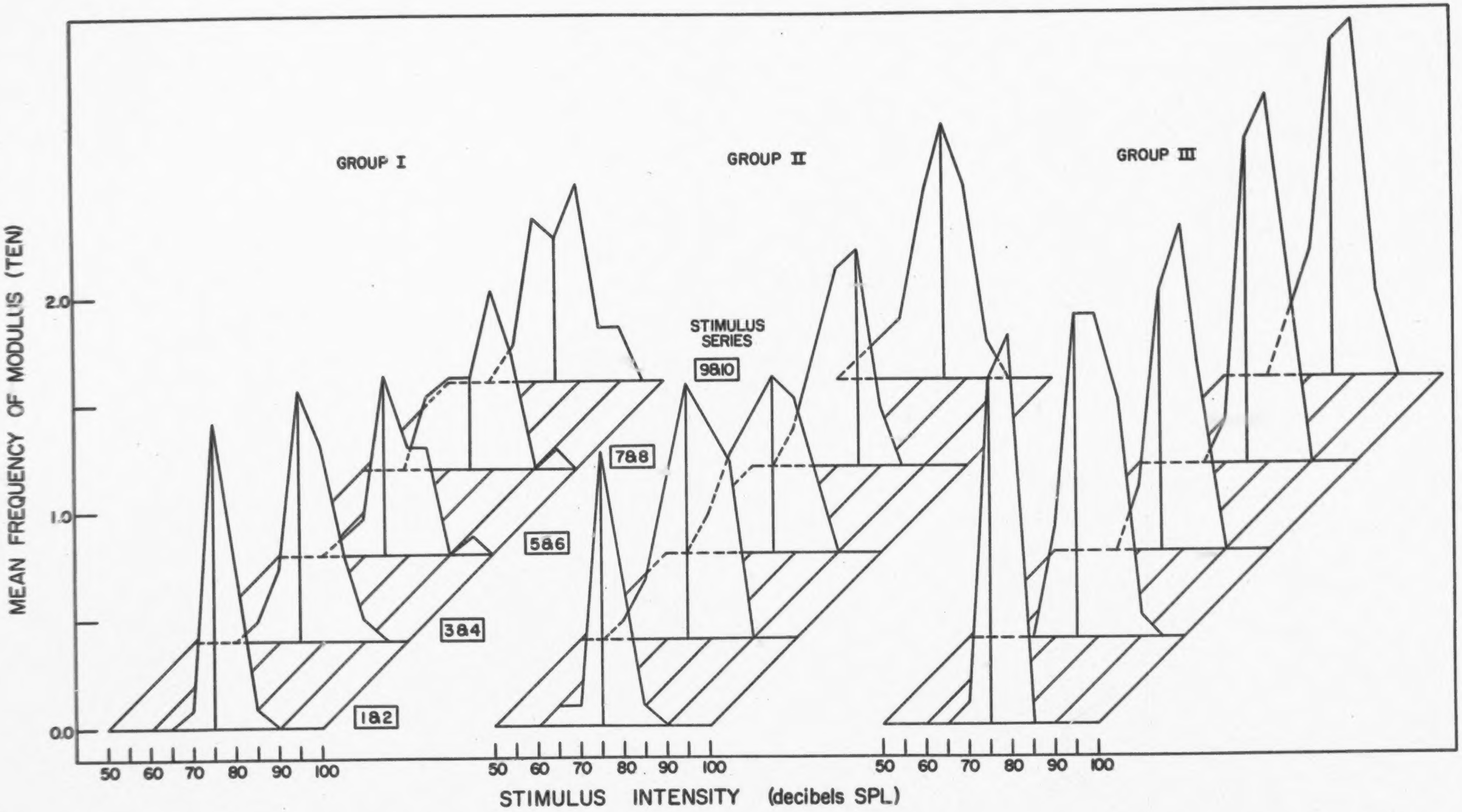
MEAN FREQUENCY OF MODULUS (TEN)
2.0
1.0
0.0
GROUP I
GROUP II
GROUP III
STIMULUS SERIES
9&10
7&8
5&6
3&4
1&2
50
60
70
80
90
100
STIMULUS INTENSITY (decibels SPL)

Fig. 3

THE JUDGMENT OF PERSONAL CHARACTERISTICS AND EMOTIONS FROM THE NONVERBAL PROPERTIES OF SPEECH

Ernest Kramer
The University of Michigan

THE JUDGMENT OF PERSONAL CHARACTERISTICS AND EMOTIONS FROM NONVERBAL PROPERTIES OF SPEECH[1]

Ernest Kramer
The University of Michigan

In a fantasy novel by the poet Robert Graves, a man of the distant future asks a twentieth-century Englishman, "Do I speak with correctitude?" "With great correctitude," he is assured, "but without the modulations of tone we English use to express, or disguise, our feelings" (Graves, 1949, p. 1). All of us not only use such modulations ourselves, but also make judgments about others' current feelings and attitudes, as well as about more stable personal characteristics, partly on the basis of how they "sound" to us. Sullivan has stated (1954, p. 7) that these "sound accompaniments suggest what is to be made of the verbal propositions stated." Whether or not we can interpret them correctly, whether or not speaker and listener would agree to their significance, these "non-verbal but nonetheless primarily vocal aspects of the exchange" (Sullivan, 1954, p. 5) play an important part in the perception of persons.

The present review of studies on the nonverbal aspects of vocal communication divides the literature into two principal categories. First, studies of the relationship between voice and judgments of relatively stable personal characteristics—those characteristics which do not fluctuate from day to day—are presented. The second part is concerned with voice and those emotional or affective variables which change over relatively short periods of time. The chief focus in both sections is on experimental studies in English; however, some foreign publications and a few theoretical and clinical papers are included.

The section on stable characteristics of the individual is subdivided into four sections. First, the major early studies of Pear (1931) and of Allport and Cantril (1934; Cantril and Allport, 1935) are presented together with their theoretical background in the work of Sapir (1927). Second are studies of the relationship between voice and physical characteristics. Included are age, appearance, birthplace and language, birth order, body type, complexion, and height. Third are studies of the relation between voice and an individual's aptitudes and interests. This section includes dominant values, intelligence, leadership, musical abilities, political preference, scholarship, and vocation. Fourth are studies of voice and personality. Included are studies of dominance, introversion-extroversion, personal adjustment within the nonclinical population, psychopathology, sociability, self-concept, and finally personality in general or global terms.

The studies of the relationship of voice to the changing emotional or affective state of the speaker are presented in chronological order of publication, except where a group of studies by a particular author clearly belongs together.

Voice and the Stable Characteristics of the Individual

The linguistic theories of Sapir (1927) provided the background for the pioneering experiments in the relation of voice to personality and emotion. Although the first of these experiments (Pear, 1931) restates Sapir's theories at some length, the particular form of the experiments does not derive directly from them. Sapir's paper (1927) gave much of the impetus for early experimental work. He divided speech into five "levels": (1) voice, (2) dynamics, (3) pronunciation, (4) vocabulary, and (5) style. The first three levels include the aspects of speech identified here as nonverbal; that is, those sounds of speech which accompany the words but do not themselves form any part of the identifying features of any particular words.

Voice. Sapir noted the absence of an adequate vocabulary for describing voice, a problem which later papers continued to demonstrate. The acoustical correlate to voice is probably to be found in the fundamental frequency and the balance of amplitudes among certain harmonics which do not carry necessary semantic information. Although Sapir did not speculate on the specific connections between voice and emotion, he felt it was "clear that the nervous processes that control voice production must share in the individual traits of the nervous organization that control personality" (p. 897). In voice, as in all the successive levels of speech, he stressed the fact that there were social as well as individual determinants.

Dynamics. This includes variations in intonation patterns, rhythm, relative continuity or grouping, and speed. The social determinants of dynamics include "the linguistically irrelevant habits of speech manipulation that are characteristic of a particular group" (p. 901), as well as elements which are part of the more recognized language of the speaker's community and which carry semantic loading.

Pronunciation. Sapir gave most attention here to the symbolic character of sounds in pronunciation. For example, in speaking to a child we are likely to change the word "tiny" [taɪni] to sound like "teeny" [tini]. There are no rules of English grammar which justify such a change, but Sapir suggested that "teeny" has a more directly symbolic character for speaking to children because of the smaller space in the mouth used for the "ee" vowel [i], which gives it the value of a gesture emphasizing the feeling of smallness. Individuals vary in both the sensitivity of their response to such factors and the extent to which such changes are adopted, consciously or unconsciously, into their own speech. Phonetic symbolism has been discussed by Brown (1958, pp. 110-154), who is more concerned with certain fixed pronunciations than with the varying pronunciations of a single word.

Vocabulary and style. These last two levels are completely outside the area of this review. Vocabulary refers specifically to those lexigraphic symbols which can be found in a dictionary, and style refers to the arrangements of these symbols into groups and the arrangements of those groups into still larger units.

Major early studies

Pear, an English psychologist, was one of the first psychologists to be stimulated by Sapir's (1927) exposition. Pear (1931) presented a summary of Sapir's paper, and then expressed skepticism about many of the popular notions concerning the characteristics expressed by voice. Errors in popular notions might arise, he felt, because the sound and look of a person are experienced as a whole. The increasing number of radios in British homes provided an opportunity to experiment with impressions based on voice alone. On three successive nights a selected passage was read over the air by three different readers each night. Report forms were requested from the listeners, and over 4000 of the reports were sent in. The form asked the listener to judge each speaker's sex, age, profession or occupation, experience in leading others, locality of birth, and other localities which affected his speech. Pear tried to avoid voices which were typical of a particular dialect or local accent. Seven of the nine readers were chosen on the basis of "achievement of *definite and recorded success* in their own calling" (p. 156-157). It was one of the other two speakers, Pear's eleven year old daughter, who provided the listeners with their only problem in judging the sex of a speaker. She was judged by 8.1% of listeners to be a boy. The judgments of the speakers' ages tended towards a median of 39 years old; speakers younger than this were judged to be older than their actual age, and older speakers were judged to be younger. The actual occupations of the speakers were a detective-sergeant, a clergyman, a buyer of ladies' tailoring, a military officer, a judge, an electrical engineer, an actor, and a private secretary. The actor and the clergyman were frequently judged correctly according to profession; 58% of the listeners who sent in report forms correctly identified the actor and 38% the clergyman. Certain errors in judging professions showed a marked consistency. For example, the detective-sergeant was judged by 50% of the respondents to have some out-of-door occupation, such as farmer or rancher. The most frequently guessed birthplaces were those who supposed dialects are frequently represented on the stage. The guesses, however, bore no relationship to the speakers' actual birthplaces. The effect on speech of a locality in which the speaker lived later in life was most recognizable for London, **Lancastershire**, and the United States. The strongest impressions of accustomed leadership were conveyed to the listeners by the actor, the judge, and the clergyman. This suggests, according to Pear, that the speaker whose voice is professionally important may have modified it, consciously or unconsciously, toward a decisive, authoritative tone.

Pear made no attempt to determine which aspects of voice were responsible for the various judgments, although he expressed the hope that future studies would investigate this. He speculated about the impressions these voices would have produced on listeners who did not know English; there is still no study which has tried to answer this question. He urged consideration of the practical consequences of the relationship between speech and personality in such fields as education, selection of teachers, and mass communication (Pear, 1932). His work was the first clear experimental demonstration of a relationship between speech and other relatively stable characteristics of an individual. It also demonstrated that there are what Pear and others have referred to as vocal stereotypes; that is, certain voices which convey a similar impression to many listeners, regardless of the correlation of that impression with other measures. The presence of vocal stereotypes remains the most frequent finding in all studies of the relationship between voice and personality.

Allport and Cantril (1934; Cantril and Allport, 1935) ran a series of experiments in which judges listened to radio voices and voices heard from behind a curtain. The voices were presented in groups of three. Judgments were requested concerning certain groups of "inner" and "outer" characteristics. Some of the characteristics were considered in a number of experiments, others in only one. Among "outer" characteristics, or physical and expressive features two experiments showed that the speakers' ages could be judged with an accuracy significantly better than would be expected from chance guessing. Only one of the four experiments which asked for judgments of the speakers' height showed results which were significantly better than chance judgments. Complexion was judged with better than chance accuracy in the one experiment which included it. The authors caution against an uncritical acceptance of so surprising a finding until more studies of it have been done. Handwriting showed no statistically significant correlation with voice in any of the five experiments in which the judges tried to match voice and handwriting. Allport and Cantril consider that the lack of significant results was due to having used judges who were untrained in handwriting analysis. Appearance of an individual in photographs was more accurately matched with his voice than was his actual appearance in person. The authors suggest that the difference is due to the necessary time lapse between hearing a speaker's voice and then having him step before the curtain, whereas with matching of voice and photographs, voice and appearance could be considered simultaneously.

Among the "inner" characteristics, or interests and traits, vocation of the speakers was judged correctly from their voices significantly more often than chance guessing would predict. Judgments of political preferences were "surprisingly successful," a result attributed to certain overly distinctive voices in each group of three speakers. Judgments of extroversion-introversion from voice correlated significantly in a positive direction with the speakers' scores on an extroversion-introversion scale by Heidbredder (not further identified) in three experiments, but gave slightly negative results in others. The correlation between the speakers' scores on the Allport A-S Reaction Study (Allport and Allport, 1928) and judgments of ascendence and submission made from their voices was significantly positive in four out of six experiments. The negative results in the other two are explained as being due to the presence in those speaker groups of an actually submissive professor who had purposely cultivated an ascendant manner for classroom purposes. Judgments of dominant values gave mixed results when compared to the speakers' scores on the Allport and Vernon Study of Values (Allport and Vernon, 1931). Summary sketches of the speakers were, on the average, more correctly matched with their voices than was any single quality.

The general findings of the study were: (1) "Many features of many personalities can be determined from voice"; (2) there is more uniformity of judgment than accuracy—stereotypes play an important part in judgments; (3) "inner" traits are judged more consistently and more correctly than are physical and expressive features; (4) judgments are influenced by heterogeneity in the judged group; and (5) the more information available about an individual, as in the summary sketches, the more accurately can his voice be matched correctly with the information. The authors concluded their study with the comment (Allport and Cantril, 1934, p. 55), "Since the criteria are imperfect, it must be borne in mind that the human voice may reveal even more concerning personality than our results indicate."

Physical characteristics

Age. Both Pear (1931) and Allport and Cantril (1934) found significant positive correlations between the age of their speakers and the estimates of age made from the speakers' voices. Both studies also found a tendency for the estimates of age to center in the thirties. Herzog (1933), in Germany, also found a significant correlation between age judgments from the voice and the true age of the speaker.

Appearance. Allport and Cantril (1934) found that judges could match a speaker's appearance with his voice with an accuracy significantly better than chance.

Birth order. Koch (1956) found differences in articulation between first- and second-born siblings. In opposite sex pairs the first born stuttered more. The reverse was true of the same sexed siblings. When siblings were very close in age, girls articulated better than boys, but at wider age spacings no sex difference was apparent. No cross-validation or attempt to postdict birth order from voice was reported.

Birthplace and language. Pear (1931) found that the birthplaces most frequently assigned by listeners to the speakers they heard were those places whose dialects, or supposed dialects, were most frequently portrayed on the stage. There was little actual correlation between the true birthplaces and those judged correct. Pear attempted to use speakers whose speech was not characteristic of a particular dialect, but there is no evidence reported as to how successful he was in this.

A person's place of birth is clearly a determinant of the language or dialect he speaks. The importance of the pitch contour of speech (the changing patterns in the fundamental frequency) in conveying impressions of accent and language was demonstrated by Cohen and Starkweather (1961). They found that English-speaking listeners could judge whether or not they were hearing a recording made from English speech, even after the recording had been passed through a low-pass filter which removed all those higher frequencies required for recognition of words (French and Steinberg, 1947). Monrad-Krohn (1947, 1957) has noted certain types of brain damage in which certain aspects of the speaker's intonation pattern, which he calls "prosody", are distorted, giving the listener the impression of a person speaking with a foreign accent.

Body type. The study by Fay and Middleton on voice and Kretschmerian body types (1940a) was one of a series of nine experiments, begun in 1939, on the relationship of voice to various stable and changing characteristics of a person. In all of the studies, the voices of the speakers were transmitted over a public address system or tape recorded, and recorded music was played in the interval between different voices. The listening judges wrote down whatever judgment the particular experiment called for during the interval of music; "there was never any silence" (1939a, p. 149). No information was reported on what music or type of music was used; it is impossible to say what effect on the judgments the music itself may have had. In their study on voice and body types (1940a), the authors took no actual morphological measurements of the speakers, but merely assigned each of them to one of the Kretschmerian types on the basis of his superficial appearance. Since actual measurements were considered a necessary step in classification by Kretschmer (1925), the speakers may not have truly represented good examples of the classification given to them. Listeners were asked to match the speakers' voices with paragraphs describing the three Kretschmerian body types. The athletic type was matched no better than if the matching had been a matter of chance alone. The pyknic and leptosomatic types were matched with only slightly better accuracy. A study, done in Germany (Bonaventura, 1935), in which voices were matched with photographs of the Kretschmerian types found the same order of relative success in matching: pyknic matched most accurately, then leptosomatic, and finally athletic.

Complexion. Allport and Cantril (1934) found their judges able to estimate speakers' complexions from their voices with statistically significant accuracy. The unexpectedness of this result caused the authors to caution strongly against accepting a general relationship without replication of the study.

Height. Herzog (1933) found that voices could be matched with the differing heights of the speakers with an accuracy better than what chance guessing would predict. Allport and Cantril (1934), however, ran four experiments in which judges attempted to match height and voice, and successful matching occurred in only one of the four.

Aptitudes and interests

Dominant values. Allport and Cantril (1934) found mixed results in a number of experiments in which listeners estimations of a speaker's dominant values were compared with the speaker's scores on the Allport-Vernon Study of Values (Allport and Vernon, 1931). The study was largely replicated by Fay and Middleton (1939a), who found a correlation of +.52 between the speaker's test placement and listeners' judgments of the dominant value types to which they belonged. Not all the Spranger value types which the Study of Values measures were equally well estimated from the speakers' voices. The types "judged most accurately in terms of mean percentage superior to chance are: political, 46 per cent; aesthetic, 29 per cent; social, 23 per cent" (p. 154).

Intelligence. Michael and Crawford (1927) had a single judge rate a number of students on various voice qualities. These ratings were compared with each student's scholarship record and with a measure of his intelligence based on a group test by Thurstone (not further identified). Low positive correlations were found between "good inflection" and both scholarship and intelligence; correlations were not improved by adding other factors to inflection. The judge, who knew some of the students prior to the experiment, did better on judging students who were unfamiliar to him. The authors ascribe this to the difficulty of judging inflection in familiar voices. Fay and Middleton (1940c) found a correlation of +.33 between estimates of intelligence from voice and speakers' I.Q.'s as measured by the Terman Group Test of Mental Ability (Terman, 1920). Positive results all came from the identification of superior I.Q.'s. Ss of below average I.Q. were, as a group, rated as higher than the average group. Fay and Middleton suggest that this finding may have been the result of one person in the average group whose voice seems to have been a remarkable stereotype of low intelligence. The authors add (p. 190), "Possibly all the ratings indicate voice stereotypes. The fact that some of them agree with the test results of intelligence may be purely coincidental."

Leadership. Pear (1931) found that his listeners gave the highest ratings of leadership to those readers whose voices were important to their professional roles: the actor, the judge, and the clergyman. He used no independent criterion for leadership. Fay and Middleton (1943) had the 15 freshman fraternity men who were their speakers rated for leadership by 10 seniors in the fraternity who had known the freshmen for six weeks. These ratings showed virtually no correlation with ratings of leadership made from the voices of the freshmen speakers. The reliability of the voice ratings was +.41. The authors feel that this degree of social agreement, in the face of no actual accuracy compared to the criterion, suggests the presence of vocal stereotypes of leadership.

Musical abilities. Ramm (1946) found that monotonism was related to below average musical abilities in 25 fifth graders. She defined monotonism as the inability to carry a tune, even though individual notes could be matched. This would seem likely to affect the intonation pattern of speech, but Ramm makes no specific mention of the children's speech. The connections between voice and music provide some of the conceptual framework for a theoretical paper on "speech melody' by Zucker (1946). He stresses the importance of speech melody in language teaching. His suggestion that one step in learning a language might be to blot out the words by some electro-acoustical device is an interesting anticipation of the later use of speech filtering in psychological studies of language.

Political preference. Allport and Cantril (1934) attribute the success of their judges in judging political preference from voice to the fact that each group of three speakers contained one person whose voice was a typical and marked stereotype of some political type.

Scholarship. There is only the study by Michael and Crawford (1927), reviewed above under the heading "Intelligence."

Vocation. Both Pear (1931) and Allport and Cantril (1934) found that listeners could judge a speaker's profession with an accuracy significantly beyond what chance alone would predict. In a later study, Fay and Middleton (1939a) found less positive results. Only the voice of a preacher was correctly identified consistently better than chance, and it was frequently mistaken for that of a lawyer.

Personality

Dominance. Allport and Cantril (1934) found a significant positive correlation between judgments of dominance made from voice and the speakers' scores on the Allport A-S Reaction Study (Allport and Allport, 1928). Eisenberg and Zalowitz (1938) used the Maslow Social Personality Inventory (Maslow, 1937) to obtain criterion scores on dominance for their 16 speakers. The judgments of dominance made from the speakers' voices show more social agreement than correctness and no relationship between agreement and correctness. The authors note that if the three or four "easiest-to-judge" voices were removed from the total of 16, judgments would be no better than one would expect on the basis of chance alone. Moore (1939) had 453 students both rate themselves and have 10 other students rate them on various personal qualities, including dominance. At least two judges, "trained in speech and working independently agreed in the classification of each voice" of the 453 (p. 33). Students who were classified as having a "breathy" voice quality were those who had ranked lowest in dominance, while those with a "nasal whine" ranked slightly higher. The study well illustrates the lack, noted by Sapir (1927), of an adequate vocabulary for describing voice. Mallory and Miller (1958) gave their Ss, all females, the Bernreuter Personality Inventory (Bernreuter, 1931) to obtain scores on submissiveness, introversion, and dominance. Judgments of dominance were then made from the Ss's readings of a standard passage. No information is reported as to who the judges were or as to what were the exact categories of judging. A "slight positive association" is reported between dominance and the voice qualities of loudness, resonance, and lower pitch. The authors state that they were testing the hypothesis that certain vocal habits are associated with particular personality traits because they are established by the same sequence of early events. It should be noted, however, that the actual study is completely ahistorical.

The Bernreuter Personality Inventory (Bernreuter, 1931), used in the study above, has been the most frequently used instrument for personality assessment in the studies reported in this review. The following cautions concerning its interpretation should be kept in mind in weighing the results of studies using it (Tyler, 1953): (1) the test is affected by the conscious as well as the unconscious set of the subject; (2) artifactual correlations within the test seem to permit clear measurement of only two characteristics, one having to do with general emotional stability and the other with sociability or self-sufficiency; and (3) adjusted and maladjusted groups show much overlap, and behavior problems cannot be identified from normals. These considerations, especially the second and third, seriously weaken the value of Bernreuter scores as criteria against which judgments based on voice may be validated. The lack of adequate independent criteria is a recurring problem throughout the studies reviewed here.

Introversion-extroversion. Moore, (1939) in the study discussed above under "Dominance," found that individuals with a "breathy" quality of voice were high in introversion, as measured by the Bernreuter Inventory (Bernreuter, 1931). Fay and Middleton (1942) found that their judges had no actual success in identifying introversion from voice, but the presence of agreement among the judges in their ratings provided further evidence for the presence of vocal stereotypes. Mallory and Miller (1958), discussed above under "Dominance," found that Bernreuter scores on introversion were related negatively to loudness, low pitch, and resonance in the voice, and unrelated to rate of speaking.

Personality adjustment. Moore (1939), in a study already noted above, found a "breathy" quality of voice positively related to neurotic tendencies as measured by the Bernreuter Inventory (Bernreuter, 1931). Duncan (1945) had his speakers rated for voice quality by fellow speech students after three weeks in class. He also obtained speakers' Social Adjustment scores on the Bell Inventory (Bell, 1934). Of the 30 descriptive voice terms used in the ratings of voice quality, 11 could be used to identify whether the speaker had been low or high in his Social Adjustment score. No cross-validation of these discriminating terms was reported. The author notes that the speakers also took the Bernreuter Inventory, but when no significant correlations between Bernreuter scores and voice ratings appeared, "the Bernreuter was excluded from further use in this study" (p. 50). Bamm (1946), in the study noted above under "Musical abilities," found that 25 fifth graders with monotonism showed inadequate social and emotional adjustment on a number of personality tests, especially the Rorschach. Unfortunately, no control group of non-monotone fifth graders was used. The lack of a control group makes the interpretation of Rorschach scores particularly tentative, since there is no adequate normative data on children's Rorschach performances as reflections of general personality adjustment. Studies on the relationship between voice and more extremely deviations in personal adjustment are reviewed below, under "Psychopathology."

Psychopathology. Although practicing clinicians have been well aware of the importance of the nonverbal aspects of voice for problems of diagnosis and therapy (Sullivan, 1954; Lacey, 1959; Shakow, 1959), few experimental studies have been done. Moskowitz (1951, 1952) studied the voices of schizophrenics, but her report on the diagnostic significance of "monotonous, ak, gloomy, and unsustained" voices is less an aid in psychodiagnosis than a reminder of Sapir's (1927) lament over the lack of an adequate language for describing voice. A study on schizophrenic children (Goldfarb, W., Braunstein, Patricia, and Lorge, I., 1956) reported that these youngsters, compared with a normal group, were ineffective in conveying mood or emotion vocally, giving the effect of either no emotion or one which "has little or no relation to the language content" (p. 549). Ostwald (1960) has made some tentative suggestions about the relationships between certain types of psychiatric patients and the spectrum analyses of their speech. His article is chiefly concerned with the techniques of speech spectrum analysis; he presents no evidence on the validity of the diagnostic impressions derived from the spectrograms. He notes that the voice records may be superimposed for comparison of different patients.

Among the publications by Moses (1941, 1942, 1954; Jones, 1942), it is his Voice of Neurosis (1954) which presented in the fullest detail the foundation and implications of his belief that "voice is the primary expression of the individual, and even through voice alone the neurotic pattern may be discovered" (p. 1). Moses is aware of the dangers of misleading vocal stereotypes. He also recognizes the need for basing judgments on different frames of reference in different social and linguistic groups. He has clearly made an attempt to set down the relevant voice variables and his method of judging them in the most objective manner possible. A large part of his work, however, remains exclusively the analysis of the single expert clinician. Some of his voice categories do seem quantifiable; for example, "range" as the range of fundamental frequencies used, and "rhythm" as a stress pattern of the changes in amplitude over time. But other categories, such as "registers" and "melism" ("the vocal means of expressing personal appeal," (p. 72), need much redefining before they can be submitted to close experimental investigation. Despite this weakness, Moses's clinical acumen and experience are important in an area marked by the inadequacy of experimental studies.

Self-concept. Moore (1939), in a study noted above, found that individuals with a "breathy" quality of voice rated themselves lower in desirable personal qualities than others rated them, while those with "harsh" and "metallic" voices rated themselves higher. Wolff (1943) had subjects rating voices on various personal characteristics. Unknown to the raters, their own voices were included. Only 10.5% recognized their own voices. The "unconscious self-judgments," as Wolff terms the ratings of the others, agreed in general with the personality ratings done of the voice by others, but they tended to judge each characteristic as more extreme or more obviously present than did others' ratings.

Sociability. Fay and Middleton (1941) obtained sociability scores on their speakers by having them take the Bernreuter Inventory (Bernreuter, 1931). The recorded voices were then presented to listeners who were asked to rate them for sociability. Each voice was presented twice, with the order of presentation changed the second time. No significant correlation was found between sociability ratings and Bernreuter scores. The reliability of listener ratings, based on the two presentations of each voice, was .40. Some voices seemed to be stereotypes of extreme sociability or unsociability.

Personality. The studies reviewed in this section are those which deal with personality in general or global terms, rather than in terms of specific aspects or traits. Allport and Cantril (1934) found that listeners were more successful in matching summary sketches of their speakers to the correct voices than they were in matching any single trait. Taylor (1934) had his speakers fill out a questionnaire of 136 items, including items from Thurstone's Personality Schedule (Thurstone and Thurstone, 1929). He then had a large number of judges, at least 20 for each voice, listen to the speakers read a standard passage. After listening, the judges filled out the same questionnaire for each speaker, as they judged the speaker to be on the basis of his voice. Detailed findings were not reported, but the general conclusions drawn were: "1. There is clearly a high degree of social agreement in judging the personality traits of people with speech as the only guide. 2. Social judgments thus based on speech bear no relationship to the judgments of the subjects themselves.... 3. There is a tendency for the auditors to be most consistent in their judgments when they agree least with the subjects' self-rating..." (p. 248). Stagner (1936) took issue with previous studies for dealing with voice as an unanalyzable whole. He had 10 speakers, reading a standard passage, rated by 25 listeners on the speech traits of voice intensity, flow of speech, poise, and clearness. The listeners also made ratings of the personality traits of aggressiveness and nervousness. Under "general impression" listeners seem to have made ratings of both general voice quality and general personality characteristics. Split-half reliabilities for all categories except aggressiveness, which was less reliably judged, were between .70 and .90. The speakers filled out the Bernreuter Personality Inventory (Bernreuter, 1931) and the Wisconsin Scale of Personality Traits (Stagner, 1937), but no consistent relationship was found between test scores and any of the listener ratings. Stagner interpreted the lack of correlation between listener ratings and Bernreuter scores to mean a lack of relationship between self-judgments of personality and social judgments based on voice. Among the personality and speech quality ratings based on voice, vocal intensity showed the least intercorrelation among vocal quality categories with the two personality trait categories of nervousness and aggression. Flow, poise, and clearness all correlated positively with aggression and negatively with nervousness. Jones (1942) gave the Rorschach test to an adolescent boy and also made a recording of the boy's voice. The test protocol was given to a well known Rorschach analyst, Piotrowski, and the voice recording was given to Moses for analysis. The two independent analyses were considered to match well with each other. Moses (1942) has listed the 21 variables he used in making his analysis. The problems involved in their validation and use by others are those discussed above under "Psychopathology," in connection with other work by Moses (1954). Brieland (1949, 1950), studying the speech of the blind, found no significant correlations between judges' ratings of effectiveness in speech and Bernreuter scores (Bernreuter, 1931) in either his blind or his sighted group.

Wolff (1943) had listeners write a free description of their impressions of the personalities of speakers whom they heard on recordings. From these descriptions, which suggested considerable communality of judgment, Wolff decided on summary terms. Listeners then heard the speakers presented in groups of three and attempted to match the voices with the summary terms. Wolff tried to estimate the actual validity of these judgments by comparing them with ratings done by personal friends of the speakers. He reports high agreement among listeners on matching voices with summary personality terms and significant agreement between these matchings and ratings by the speakers' friends. The details of the findings are not reported. Starkweather (1955b, 1956b) studied vocal differences between a normal group and a group which had high scores on a personality test (Harris, 1953) which distinguishes hypertensives from normals. He predicted that individuals with the hypertensive personality syndrome would show greater incongruence between the verbal and the nonverbal aspects of their speech than would normals. The measure of incongruence was the discrepancy between judges' ratings of the emotional content of typescripts of what the subjects had said and their ratings of recordings of the subjects' speech from which the verbal aspects had been removed. The hypothesis was not supported, but the study is interesting because of the technique used for removing verbal content. The speech samples were passed through a low-pass filter which held back those higher frequencies of sound upon which word recognition depends (French and Steinberg, 1947; Licklider and Miller, 1951). The characteristic personal tone quality of the voice is also altered by the filtering, but many of the nonverbal aspects, such as stress patterns and intonation patterns based on changes in the fundamental frequency, still remain. Starkweather used recordings of role-playing sessions which had been recorded on

Gray Audograph Equipment at Berkeley's Institute for Personality Assessment and Research. It seems quite possible that this low-fidelity sound equipment cut out more frequencies than merely those above 450 cps which Starkweather meant to eliminate; it also may well have introduced some distortions into the recording. Despite these inadequacies in the recording equipment, the judges showed significant agreement in rating the emotional content of the filtered speech.

Voice and Changing Emotional States

The effect of emotions on the throat and breathing muscles involved in voice production was noted by speech teachers (Blanton, 1915) even before any psychological studies of voice and emotion were done. Lynch (1934) considered fundamental frequency to be one of the parameters of speech most likely to be affected by emotionally mediated tension. In his experiment, both trained and untrained readers read not only factual material, but also dramatic material calling for grief and anger. Trained readers showed more variation in the fundamental frequency of their voices between the different types of reading; they also showed more variety in fundamental frequency among themselves than did untrained readers. For both groups of readers the average pitch level was highest for anger, next highest for grief, and lowest for factual material. The pitch range in the readings was greatest for anger, next widest for factual, and narrowest for material calling for grief. Skinner (1935) tried to eliminate the presence of verbal content by having his subjects say merely "ah." They first read a passage of emotional literature and listened to selected music in order to put themselves into a happy or sad emotional state. No independent measure was included in the study of how successful this emotion-inducing procedure proved to be. Skinner found that the ah's of happiness showed higher pitch and greater force than those of sadness. Ortleb (1937) had his subjects read emotional literature aloud. He found that pitch, intensity, and duration tend to rise together in emphasized syllables. Fairbanks and Pronovost (1939; Fairbanks, 1940) had actors read five passages, each marked by a different emotion. Listeners heard only a set of sentences that was common to the five passages, and were asked to identify the emotional tone of the entire original passage from the sound of these excerpts. Some of the actors seemed to provide much clearer vocal differentiation of emotion than did others. The authors found measurable pitch differences among the different emotions, using average measures from the different readings. This and all other studies using trained actors must, however, be interpreted with caution. Stage speech seems to have markedly different qualities from normal speech (Cowan, 1936). Furthermore, an actor is likely to portray just that stereotype of emotion which listeners from the same social-cultural milieu would find easiest to recognize.

Dusenbury and Knower (1939) asked a group of speech students and instructors to "try to feel the designated emotional state and to use a tonal code which would indicate their feelings" (p. 67) while reciting the letters A through K. Eleven emotions were designated. Twenty-two such recitations were recorded, and eight of these sets were selected on the basis of pretests to be matched by listeners with a list of emotions. All recordings were matched with the emotion which the speaker had tried to represent with significantly greater than chance accuracy. Another group of listeners heard only part of each A through K recitation, and they did significantly less well at matching them with the correct emotions. This study, like similar ones described below, may merely have measured the ability of an individual to communicate a shared vocal stereotype to his listeners. The study does not consider whether these individuals, or any others, would use these particular "tonal codes" when experiencing these emotions in real life situations. Knower (1941) had speakers both speak and whisper the letters A through K in terms of a designated emotional state. Different groups of listeners tried to match emotions with these recordings played forwards, and other groups did their matching with the recordings played backwards. Knower felt that whispered speech would eliminate the effects of "tone," since the fundamental frequency of the voice is not present in a whisper. Playing the recordings backwards was intended to investigate the effects of "pattern" on the emotional expression. All conditions are reported as giving better than chance recognition of the emotions the speakers were trying to express. Decreasingly successful results were found for voiced speech played forwards, whispered speech forwards, voiced backwards, and whispered backwards.

Fay and Middleton (1940b) tried to determine whether listeners could distinguish whether or not a speaker was rested or tired by the sound of his voice. The "rested" speakers had their normal amounts of sleep, while the "tired" group had gone without sleep for 30 hours. None of the speakers had any speech defects, "nor, in the opinion of the writers, did any of the speakers possess voices noticeably lacking in vitality" (p. 646). This opinion seems not to have corresponded with the perceptions of the listeners. The accuracy of the listeners in assigning speakers to the correct group, rested or tired, was less than would be predicted by chance. The authors feel that "the existence of stereotyped tired and rested voices" (p. 649) was probably the reason for the worse than chance results. In another study, Fay and Middleton (1941b) asked listeners to judge whether a speaker was telling the truth or lying. Lying was identified with "an accuracy slightly exceeding chance" (p. 215), while truth-telling seemed to have no distinguishing characteristics which permitted better than random guessing.

Fairbanks and Hoagland (1941) selected a passage of prose which was subject to different interpretations. They had six different amateur actors each read the passage with five different simulated emotional states. Listeners could differentiate representations of anger, fear, and indifference as a group from representations of contempt and grief. Anger, fear, and indifference were all characterized by a rapid rate of speaking, with short phonations and short pauses, but they could not be distinguished from each other by these measures. The representations of contempt and grief were both characterized by a slow rate of speaking. Both phonation and pauses were equally prolonged in passages read with contempt. The relatively slow rate of speaking which marked representations of grief was almost entirely due to prolongation of pauses, particularly between phrases.

Brody (1943) called attention to subtle variations in patients' voices during the course of psychoanalysis. He presented several cases in which vocal changes seemed to mark major emotional stages in therapy. He regards vocal expression as a relatively safe way to act out hostile feelings during analysis.

Baker and Harris (1949) had their subjects take a word intelligibility test, speaking words aloud, under conditions of stress and no stress. In the stress condition the speakers were threatened with the possibility of electric shock (never actually given). The subjects later took the Rorschach Test, and these results were compared with their articulation scores and their speech intensity (average speech power) under stress. Unfortunately, the scoring system used for the Rorschach was, although based on other systems, highly adapted for this study, making results hard to generalize. Form level, which here seems to be a measure of the ability to see the stimuli as most others see them, was positively related to variability of intensity. One possible interpretation is that form level, which has been interpreted as a sign of ego strength (Klopfer, B., Ainsworth, Mary, Klopfer, W. G., and Holt, R. R., 1954), reflected an ability and freedom to vary performance under varying conditions of stress; no instruction to keep intensity constant was given to the subjects.

Thompson and Bradway (1950) had two psychologists act out a therapeutic interview in which they actually spoke only numbers, although with the inflections which a genuine exchange between patient and therapist might have had. The two participants each listened separately to recordings of the session and made statements about the "affective interchange." The authors report that the statements of each psychologist were significantly correlated with those of the other. The technique is recommended for use in teaching psychotherapy, based on the authors' assumption that, "when in a content-free interview, one takes the role of a therapist, he feels like a therapist. When he takes a patient role, he feels like a patient" (p. 323). Pfaff (1954) had an "experienced speaker" use numerals to express a variety of emotions. Various groups of listeners tried to identify the emotions portrayed. The listeners were college students with speech problems, college students without speech problems, college majors in oral interpretation and in mathematics, and junior high school students of above and below average socio-economic status. All groups did better than chance at guessing the emotions. As a whole, those college students without speech problems did better than those with speech problems. Junior high school students of lower socio-economic status did least well of any group at identifying the emotions, while college oral interpretation students did best. A partial interpretation of the results may be that the "experienced speaker" drew from the same stock of stereotypes and stage techniques with which the college oral interpretation students were most familiar. The low ranking of the junior high school students of low socio-economic status suggests the hypothesis that the "tonal affect language" may be different, at least to some extent, for different classes in a society.

Soskin (1953) described vocal communication in terms of two channels in a paper which dealt principally with the implications of these channels for psychotherapy. "Semantic information" is carried in the channel consisting of the articulated patterns of sounds which we recognize as words and sentences.

"Affective information" is carried in the channel bearing the changing, nonverbal features of the voice. This affect channel is the first one recognized by the infant. Later, as the child learns words, he sometimes finds himself in conflict over contradictory messages: words spoken with an emotional message that belies their semantic content. In adult life, we ordinarily expect a listener to focus most of his attention on the semantic content, and we object if he ignores this in order to focus primarily on the affective message. In psychotherapy, however, the therapist may choose to concentrate on the affective channel. This channel is less consciously controlled than is the semantic channel. One goal of the psychotherapist may be to enable the patient to recognize for himself the nature of the affective message he is communicating. Soskin and Kauffman (1961; Kauffman, 1954) studied the interactions of these channels through a technique which permitted some separation of the two. They passed speech samples through a low-pass filter which sharply attenuated frequencies above 450 cps. With the higher frequencies removed, speech intelligibility remains only for some common prepositions, articles, and conjunctions; the nouns and verbs which make continuing semantic content clear can no longer be recognized (Fletcher, 1953; French and Steinberg, 1947). Early studies with this technique made it clear that the remaining frequencies still carry much affective information (Soskin and Kauffman, 1961). Fifteen voice samples and a list of emotions for categorizing them were presented to two groups of listeners. One group heard the samples after they had been passed through the filter, the other group heard the unfiltered speech. The emotional category most frequently chosen for each sample was generally the same for the two groups. The authors next presented filtered speech samples to a group of listeners who had been given a special scheme for categorizing the emotions. The samples were first judged for the major emotional states involved, then for subdivisions of these states, and finally for still finer subcategories. The listeners generally agreed significantly with each other in their use of the first two levels, but not at the third level. The experimenters note that the filtering technique effectively eliminates the semantic channel, but unfortunately also eliminates that part of the affective channel message which they feel resides in the middle frequencies of speech.

Kauffman (1954) had a professional actor record two readings of a series of short speeches. In one reading the actor read with an emotional expression which was appropriate to the words of each speech, while in the other reading he used an expression which was highly incongruent with the verbal content. The recordings were passed through a low-pass filter to remove the semantic content. One group of listeners judged the second series of speeches for incongruity by comparing the filtered recordings with typescripts of the speeches. Separate groups rated the typescripts alone, and the full range and filtered recordings. The rating scheme was similar to that described above (Soskin and Kauffman, 1961), but Kauffman also classified the "meanings" of the rating categories into two main divisions: (1) expressive, "affect meanings relevant to the psychological state of the speaker," and (2) "manipulative...meanings relevant to the purposive behavior of the speaker." He found that both the vocal and verbal channels, corresponding to the affective and semantic channels of Soskin (1953), carry information about both the expressive and manipulative meanings in speech. There is, however, a tendency for the expressive function to be performed by the vocal channel and the manipulative by the verbal. Incongruence between vocal and verbal channels was reflected in greater heterogeneity of judgments, particularly in the judging of expressive meanings by those who heard only the filtered recordings. Heterogeneity of judgments was assumed to be a measure of ambiguity. There was, then, a consistent negative correlation between the degree of congruence of the vocal and verbal channels and the amount of ambiguity.

Starkweather (1955a, 1956a) sampled recordings of the 1954 Army-McCarthy hearings for three excerpts each of the voices of Senator McCarthy and Mr. Welch. The excerpts were chosen to fit the categories: matter-of-fact, challenging, and indignant. Word-free recordings were prepared by passing the excerpts through a low-pass filter which sharply attenuated frequencies above 300 cps. These filtered samples were presented twice, in counter-balanced order, to 12 clinical psychologists who rated them for which of the three context categories they best fitted, for their degree of pleasantness and unpleasantness, and for the amount of emotion present. There was significant interjudge agreement, although the judges themselves insisted that they had no confidence in their own ratings.

The judged amounts of pleasantness and over-all emotion present tended to increase as more judgments were made. The raters were then given a normal, unfiltered presentation of the excerpts and asked to place them again into the most appropriate context category. A comparison of the categories assigned to the filtered and unfiltered recordings indicates that Mr. Welch's voice was judged appropriate to the verbal content, while the Senator's voice was judged to be without variation.

Black and Dreher (1955) reported a series of four experiments relevant to the nonverbal aspects of speech. In the first, listeners were given recordings, played at various speeds, of unfamiliar voices reading factual material. The listeners' task was to restore the voice recordings to their original speeds by adjusting the turntable speed. Their success in doing so is regarded by the authors as demonstrating that there is some combined pattern of pitch, rate, and timbre which constitutes for most listeners the sound of a "normal" voice. In the second study, readers recorded a factual passage before and after their pupils had been dilated by a drug. The authors give no information as to what drug was used or what its effect on other bodily functions may have been. The post-dilation readings were selected so as not to include any which were obviously heavily distorted by reading errors and hesitations as a function of increasing dilation. Listeners were asked to judge whether readings sounded "certain" or "uncertain." Post-dilation readings were judged significantly more uncertain. Listeners also judged speakers' voices to be more tired and less alert after dilation than before. Fay and Middleton (1940b) had failed to find differences between rested and tired speakers where separate groups were used for each condition, but differences were noted in the present study where the same group of speakers were heard under two conditions.

Hargreaves and Starkweather (1961) have demonstrated that certain drugs have their own marked effects on vocal behavior. In their third study, Black and Dreher asked inexperienced readers to simulate certainty or uncertainty in reading a list of five-syllable phrases. Listeners were able to distinguish between the two types of reading with high reliability. For the fourth study, the phrase "some of them like to hurry" was imbedded in different contexts, calling for characterizations of a police sergeant, a business man, and a funeral director. The readers were 12 unselected NROTC personnel. Listeners with lists of the three characterizations were able to match hearings of the key phrase with the appropriate characterization for nine of the 12 readers. Measurements of average fundamental frequency, changes in fundamental frequency, duration of the reading, and sound pressure level were made of the recordings. No single variable differentiated consistently among the three types of characterizations. The authors suggest that the tendencies and interactions which may possibly have provided the cues for identification were: policeman—loud, low pitch; business man—slow, variable pitch; funeral director—fast, soft, monopitch.

Goldman-Eisler measured different aspects of rate of speech and breathing rate in a series of studies on noncontent aspects of speech during psychotherapeutic interviews (1955, 1956a, 1956b). The most fully reported of these studies (1956b) focuses on rate of respiration (RR) and expulsion rate of syllables (ER). RR was hypothesized to be an indicator of strength of effect, and ER was presumed to indicate ease and spontaneity of expression of affect. A "ventilation index" of RR/ER was also considered an important measure: "... high values belong to content implying free-flowing or outgoing affect...and the low indices...belong to topics of restricted emotionality, which implies tension states..." (p. 48). These measures were taken on eight patients during psychotherapy interviews and were compared with the emotions presented in the interviews. For five of the patients the emotional classification of the content was apparently considered self-evident; a single psychiatrist rated the content for each of the remaining three patients. The content, thus classified, supported the hypotheses concerning ER, RR, and RR/ER. With this kind of "validation" the study seems more a source than a test of hypotheses.

Mahl (1956, 1959) has been concerned with the role of silences and disturbances in speech as indicators of changing emotional states in psychotherapy. At times he has referred to these as "expressive aspects of...speech" (1959), but they have chiefly been treated as disruptions in the speech process rather than as part of the simultaneous nonverbal accompaniment to spoken words which is the center of focus in the present review of studies. Mahl initially concentrated on two measures (1956), the "Speech-Disturbance Ratio" and the "Patient-Silence Ratio." The first of these is the total number of such disturbances as ah's, corrections in sentences, stutters, intruding incoherent sounds, and tongue slips, divided by the total number of words the patient speaks; the second is the ratio of seconds of silence to total number of seconds available to the patient in which he might speak. These measures were validated against judgments of emotional change made from typescripts which the secretary had cleared of speech disturbances. The typescripts were divided into "motivational phases," and the change from phase to phase compared with changing values for the two ratios. Mahl concluded that "the two measures discriminate 'something' between-sessions and within-sessions for a given patient" (1956, p. 14). Krause (1961) has shown that Mahl's measures are highly similar to the measures of speech disruption used by Dibner (1956) as indicators of anxiety. Krause and Pilisuk (1961), using measures from both Mahl and Dibner, found that "intrusive nonverbal sounds, mainly laughs and sighs" were the best indicators of transitory anxiety.

Ochai and Fukumura (1957) found that the quality they term "naturalness" in voice "is distributed in the upper and lower regions (of the speech spectrum) almost uniformly," as are the "timbre nuance shades in personal voice..." (p. 393). This might serve as a cautioning reminder of how much of the nonverbal, yet personal, attributes of voice are lost in studies where verbal content is removed by removing the higher frequencies of speech.

Effects of praise and blame on speech physiology are noted by Malmo, Boag, and Smith (1957). Following praise, the tension of the subject's speech muscles, as measured by electromyograph, falls off rapidly, as contrasted with the sustained tension following adverse criticism. Measurements on the psychologist who was offering the praise or blame showed the same muscle tension pattern in him. In another part of the study, involving a series of interviews over a number of days, the authors found that the subjects had cardiac changes which correlated significantly at the 1% level with the interviewer's "good" or "bad" days as noted in his diary of his own feelings and moods. Yet the only variable in the actual interview situation which seemed to vary with the cardiac changes and the "good" and "bad" days was that on the examiner's very worst days his voice "may have been higher in pitch and smoother in texture" (Lacey, 1959).

Pittinger and Smith (1957) have suggested that certain approaches and classification schemes drawn from linguistics might contribute to psychological and psychiatric explorations of voice. Their "vocal qualifiers" include six categories: (1) intensity, (2) over-all pitch range, (3) pitch intervals between the four pitch levels of the intonation pattern (4) degree of tension or laxness of the vocal organs, and (5) tempo for the sequential march of words within the context. These categories overlap with those used in a study by Eldred and Price (1958), which also drew on linguistic schemata. The categories of alterations of pitch, of volume, and of rate are used in microlinguistics, and Eldred and Price's fourth category, break-up, is related to the linguistic classification of juncture disruptions. Four listeners judged tape recordings from various stages of intensive psychotherapy with a single patient. The same judges noted the linguistic categories and the emotional changes in the recordings. "Feelings of anger" seemed to be associated with excess loudness, fastness, and high pitch; while "depression" was associated with excess softness, slowness, and low pitch. "Anxiety and suppression of feeling" appeared to result in an increase of break-up. No cross-validation of these relationships was reported. McQuown (1957) used similar linguistic categories in a thorough analysis of a recorded sample from one psychiatric interview. He noted linguistic features which characterized the participants, and he interpreted these in terms of affect and affective communication. No independent measure of the affective aspects of the interview is reported.

Greenfield (1958) produced conflict in response tendencies in a paired-associates learning situation. He found that conflict produced significant dispersions of the formants of speech. Scott (1958) discussed the importance of vocal noises in psychoanalysis. Scott, like Mahl (1956), is concerned chiefly with disruptions in speech; his statement that "noise links us as adults to infancy" (p. 111) is reminiscent of Soskin's (1953) emphasis on the nonverbal aspects of speech in early childhood. Experimental investigation of developmental changes in production and perception of nonverbal vocal communication still remains to be done. Diehl, White, and Burk (1959) gave the Taylor Manifest Anxiety Scale (Taylor, 1953) to 178 seminary students. The students read passages from Matthew 5 aloud, and the authors of the study classified the voices into normal, hoarse-breathy, harsh, and nasal. Using the test scores as criteria, the students with hoarse-breathy voices were found to be significantly more anxious than those with either normal or harsh voices.

Davitz and Davitz (1959a) gave eight speakers a list of ten feelings, together with a paragraph describing a situation in which each would be likely to occur. The speakers were asked to express each feeling by reciting the alphabet with an appropriate expression. Thirty listeners tried to identify the feelings. All feelings were identified more consistently than chance alone would predict; the most frequently correctly identified were anger, nervousness, sadness, and happiness. Accuracy of the judges varied, but all did better than chance expectancy. There were also differences in the degree to which recitations by different speakers were correctly identified, but the presentations of all the speakers were identified at better than the .01 level of significance. Some errors in guessing showed significant consistency. When fear was mistakenly identified, it was most commonly taken for nervousness; love was most commonly misidentified as sadness, and pride for satisfaction. In a second study (Davitz and Davitz, 1959b), the authors selected from pretests two speakers who were particularly successful at communicating feelings through reciting the alphabet. These speakers each used the alphabet to express 50 different feelings. Thirty judges tried to match each recitation with the correct feeling. Ten judges rated each feeling from the list of 50, checking each feeling on the list which was similar to it. A similarity score was based on the number of times a feeling was noted as being similar to any other feeling. Strength scores were derived from the ratings of 15 judges who scaled each feeling on a six-point scale, from "very strong" to "very weak." Activity scores were obtained by having another group of judges rate the feelings on a scale running from "very active" to "very passive." A third group of judges provided the data for valence scores by rating the feelings in a scale running from "very good" to "very bad." Findings were: (1) accuracy of identification of feelings was correlated -.29 with similarity scores (significant at the .025 level); (2) the degree to which one feeling is mistaken for another is related to the subjective similarity of the two; (3) for pairs of similar feelings, the stronger tends to be communicated more accurately; and (4) no significant relationships appeared involving the activity or the valence scores. The authors noted that "since the relationships found were not high, the greater part of the variance in accuracy of communication is unaccounted for" (p. 116).

In a symposium on psychotherapy research (Rubinstein and Parloff, 1959), Lacey and Shakow drew attention to the importance of voice variables and the need for further study of them. Lacey pointed to the paper by Malmo et al. (1957), discussed above, as an example of the subtle changes in voice which might effect physiology and emotion. Shakow included "vocalization quality" as one of the classes of communication required to encompass all the important data from the psychotherapy process. He noted the heavy emphasis on typescripts in therapy research and asked, "in the process of objectifying through part analysis of the data, what...gets lost by considering only the contentual material without the associated vocal qualifiers?" (Shakow, 1959, p. 112).

Pollack, Rubinstein, and Horowitz (1960a, 1960b) had four speakers read two sentences of neutral content in 16 different "modes," such as pedantry, boredom, disbelief. Listeners matched the readings with a list of the modes. Those modes easily confused with a more frequently chosen one were dropped, and the experiment was repeated with only eight. Listener recognition was more accurate with the reduced number of modes. The authors had eight short, neurtral sentences read in various modes under increasing signal/noise ratios. They found that recognition of the modes held up better under noise than did recognition of the particular sentences. Mode recognition was also possible with significant accuracy when the sentences were whispered. Apparently information other than pitch enters importantly into the recognition process, since the fundamental frequency of the voice is absent in whispered speech. The effects of temporal sampling were also explored; some recognition of the modes was still possible with extremely short samples and with sections of the samples removed at periodic intervals.

Dittman and Wynne (1961) took excerpts from recorded psychiatric interviews and radio conversations and tried to code them for linguistic and paralinguistic categories. The linguistic phenomena—juncture, stress, and pitch—showed high inter-coder reliability, but they showed no relationship to the emotional state of the speaker. The paralinguistic phenomena—vocalizations, voice quality, and voice set—could not be reliably coded. The authors feel that "global judgments of emotionality in speech where cognitive messages are filtered out electronically... may prove useful in the analysis of interviews long before the elements which form the basis for those judgments are understood" (p. 204).

Starkweather (1961) has noted some approaches to duration and other physical aspects of the speech signals which have not yet been directly used in voice and emotion studies, but which may offer better ways of quantifying some of the dimensions of speech which change with changing emotions. Hargreaves and Starkweather (1961a) present a case study where judges were able to use certain aspects of speech spectrograph records to identify changes in a patient's vocal behavior which had been considered significant by her therapist. The validity of the "machine method" of identifying emotionally significant vocal changes still rests on the validity of the skilled listener who sets the criterion dimensions for it, but the authors feel that the method offers a great saving in effort over having a skilled listener consider separately every section of vocal behavior in an interview. Using set aspects of the spectrogram also avoids the effects of fatigue and the learning of wrong cues which might mar the judgments of the skilled listener alone. The four measures which the authors used were average intensity, frequency of the highest spectrogram peak, frequency of the second highest peak, and flatness—or the number of channels (third octave filters were used) which were 2/3 as high as the highest peak. But they point out that much more information is present in the spectrogram, and a different selection of dimensions might have served as well for finding correlates to emotional changes in the patient. It may be necessary to use different dimensions for different individuals, as Krause (1961) has found different behavioral measures of vocal behavior to be important for different subjects in identifying anxiety.

Summary and Conclusions

Common observation and popular opinion set forth the hypothesis that much more communicative information is carried by speech than is contained merely in the particular words spoken. Clinical psychologists and psychiatrists have frequently shown awareness of the importance of these nonverbal, or nonlexical, aspects of speech. Speech spectrograms, with their harmonic analysis of sound, offer a potential tool for the scientific quantification and study of this nonverbal vocal communication. Harmonic analysis could also provide an objectified record of studies involving grosser behavioral measures, thus making possible more exact replications and better comparability of such experiments. For practical purposes, a manageable number of variables would have to be selected from the large amount of information potentially present in a speech spectrogram.

Few of the actual studies of the nonverbal aspects of speech have used acoustical analysis. The grosser measures and impressionistic language of most studies frequently make the findings ambiguous to interpret and difficult to compare with each other. As a whole, there is support for the popular notion that one can tell something about a person and his feelings from the sound of his voice and the manner of his speaking. Positive results have occurred slightly more often in experiments asking for judgments of a person's changing emotions than in those which asked for judgments of stabler aspects of his personality. In neither case have results been as clear-cut as popular, unsophisticated opinion suggests.

The most consistent single finding is that agreement among listeners is greater than the correlation of their judgments with an independent criterion. The various studies have generally attributed this to the existence of stereotyped voices and vocal mannerisms to which most people give a common interpretation, independent of the correctness of that interpretation. The ratings have been considered to have greater "reliability" than "validity." What is it that the judges are reliably judging? Only Pear (1931) gives any evidence for the origin of the stereotypes. His data suggest that some of them are due to conventionalized theatrical portrayals, but the source of others is still open to question, and the origin of the theatrical conventions is also unknown. Examination of the criteria of most studies suggests part of the answer. These validity criteria, such as the Bernreuter Personality Inventory (Bernreuter, 1931), are often highly imperfect measures themselves of those traits they are used to validate (McKelvey, 1953; Tyler, 1953). A trait such as "introversion" might be as validly measured by judgments of voice as by a test scale, yet each measure could cover different portions of the total variance and thus show no correlation with the other. Better measures of personality and emotion must be used to evaluate the validity of judgments based on voice and to explain the reliability of those judgments.

There are no reported studies on how nonverbal vocal communication is learned. Apparently the developmental hypotheses advanced by some psychoanalysts (Isakower, 1939; Glauber, 1944) have been of little heuristic value to the experimentalists. There is a paucity of studies of voice and psychopathology, which may be due in part to the lack of adequate diagnostic tests and criteria. Studies of voice and changing emotions have suffered from the lack of commonly agreed upon and identifiable categories for classifying emotions.

Most of the speech samples used have been laboratory products. The speaker has read a standard passage into a microphone or has recited numerals or letters in a manner aimed at portraying a given emotion. There is a need for gathering samples of spontaneous emotional speech and treating it in some manner which separates the verbal and nonverbal elements, so that the communicative value of the latter may be judged. Filtering out the high frequencies which permit identification of words has been the most successful solution so far, but much of the nonverbal information is also lost in this process.

Attention has been given to differences among speakers, but individual differences among listeners have been neglected. Four sources of differences are:

(1) Personality variables. In addition to the effect these may have on the ability to make general personality judgments from voice, the literature on perceptual defense and need-motivated perception suggests that the listener's motivational-need structure may strongly influence his judgments of particular characteristics in others.

(2) Developmental variables. Theorizing has suggested that children may be relatively more sensitive than adults to the nonverbal aspects of speech (Soskin, 1953; Soskin and Kauffman, 1961), but no empirical evidence has been reported.

(3) Psychophysical variables. How do individual differences in acuity to the various dimensions of sound affect listeners' perceptions of a speaker's personality and emotions?

(4) Cultural-linguistic variables. In what way do the nonverbal cues in speech vary from one language group to another? It may be that each language has not only a lexigraphic vocabulary and grammar, but an equally real, though uncodified, grammar of emotions in speech.

In 1942 Sanford noted that common experience seems to accept the existence of connections between voice and personality, and even if "the analytic-experimental approach... reveals no relationship, we should be forced to conclude that it may be the fault of the approach" (Sanford, 1942, p. 838). Diehl (1960) feels "it is logical to assume that the vocal mechanism... should be responsive to all emotional states" (p. 175). The "analytic-experimental approach" has, by now, verified that a relationship does exist. Many details of that relationship still remain to be explored.

REFERENCES

Allport, G. & Allport, F. *A-S reaction study*. Boston: Houghton Mifflin, 1928.

Allport, G. & Cantril, H. Judging personality from the voice. *J. Soc. Psychol.*, 1934, 5, 37-55.

Allport, G. & Vernon, P. *Study of values: a scale for measuring the dominant interests in personality*. Boston: Houghton Mifflin, 1931.

Baker, L. & Harris, J. The validation of Rorschach test results against laboratory behavior. *J. clin. Psychol.*, 1949, 5, 161-164.

Bell, H. M. *Adjustment inventory*. Stanford, Calif.: Stanford Univ., 1934.

Bernreuter, R. *Personality inventory*. Stanford, Calif.: Stanford Univ., 1931.

Black, J. W. & Dreher, J. J. *Non-verbal messages in voice communication*. USN Sch. Aviat. Med. Res. Rep., 1955, No. NM001 104 500, 45.

Blanton, S. The voice and the emotions. *Quart. J. pub. Spkg.*, 1915, 1, 154-172.

Bonaventura, M. Ausdruck der Persönlichkeit in der Sprechstimme und in Photogramm. *Arch. ges. Psychol.*, 1935, 94, 501-570.

REFERENCES (Continued)

Brieland, D. M. Speech and personality factors in the education of the blind. *Amer. Psychologist*, 1949, 4, 291-292. (Abstract)

Brieland, D. M. A comparative study of the speech of blind and sighted children. *Speech Monogr.*, 1950, 17, 98-103.

Brody, M. W. Neurotic manifestations of the voice. *Psychoanal. Quart.*, 1943, 12, 371-380.

Brown, R. *Words and things*. Glencoe, Ill.: Free Press, 1958. Pp. 110-154.

Cantril, H. & Allport, G. *The psychology of radio*. New York: Harper, 1935.

Cohen, A. & Starkweather, J. Vocal cues to language identification. *Amer. J. Psychol.*, 1961, 74, 90-93.

Cowan, M. The pitch and intensity characteristics of stage speech. *Arch. Speech*, Supplement, 1936.

Davitz, J. R. & Davitz, Lois J. The communication feelings by content-free speech. *J. Communication*, 1959, 9, 6-13. (a)

Davitz, J. R. & Davitz, Lois J. Correlates of accuracy in the communication of feelings. *J. Communication*, 1959, 9, 110-117. (b)

Dibner, A. S. Cue-counting: a measure of anxiety in interviews. *J. consult. Psychol.*, 1956, 20, 475-478.

REFERENCES (Continued)

Diehl, C. F. Voice and personality, an evaluation. In Barbara, D. A. (ed). Psychological and psychiatric aspects of speech and hearing. Springfield: Charles C. Thomas, 1960. Pp. 171-203.

Diehl, C. F., White, R. & Burk, K. Voice quality and anxiety. J. speech and hearing Res., 1959, 2, 282-285.

Dittman, A. T. & Wynne, L. C. Linguistic techniques and the analysis of emotionality in interviews. J. abnorm. soc. Psychol., 1961, 63, 201-204.

Downey, J. E. The will-temperament and its testing. Yonkers: World Book, 1923.

Dreher, J. J. & Bragg, V. C. Evaluation of voice normality. Speech Monogr., 1953, 20, 74-78.

Duncan, M. H. An experimental study of some of the relationships between voice and personality among students of speech. Speech Monogr., 1945, 12, 47-61.

Dusenbury, D. & Knower, F. H. Experimental studies of the symbolism of action and voice: II. A study of the specificity of meaning in abstract tonal symbols. Quart. J. Speech, 1939, 25, 67-75.

Eisenberg, P. & Zalowitz, E. Judgments of dominance-feeling from phonograph records of voice. J. appl. Psychol., 1938, 22, 620-631.

REFERENCES (Continued)

Eldred, S. H. & Price, D. B. A linguistic evaluation of feeling states in psychotherapy. *Psychiatry*, 1958, 21, 115-121.

English, H. & English, Ava. *A comprehensive dictionary of psychological and psychoanalytic terms*. Longmans, Green: New York, 1958.

Fairbanks, G. Recent experimental investigations of vocal pitch in voice. *J. Acoust. Soc. Amer.*, 1940, 11, 457-466.

Fairbanks, G. & Hoaglin, L. W. An experimental study of the durational characteristics of the voice during the expression of emotion. *Speech Monogr.*, 1941, 8, 85-90.

Fairbanks, G. & Pronovost, W. An experimental study of the pitch characteristics of the voice during the expression of emotions. *Speech Monogr.*, 1939, 6, 87-104.

Fay, P. J. & Middleton, W. C. Judgment of Spranger personality types from the voice as transmitted over a public address system. *Character & Pers.*, 1939, 8, 144-155. (a)

Fay, P. J. & Middleton, W. C. Judgment of occupation from the voice as transmitted over a public address system and over a radio. *J. appl. Psychol.*, 1939, 23, 586-601. (b)

REFERENCES (Continued)

Fay, P. J. & Middleton, W. C. Judgment of Kretschmerian body types from the voice as transmitted over a public address system. J. soc. Psychol., 1940, 12, 151-162. (a)

Fay, P. J. & Middleton, W. C. The ability to judge the rested or tired condition of a speaker from his voice as transmitted over a public address system. J. appl. Psychol., 1940, 24, 645-650. (b)

Fay, P. J. & Middleton, W. C. Judgment of intelligence from the voice as transmitted over a public address system. Sociometry, 1940, 3, 186-191.(c)

Fay, P. J. & Middleton, W. C. The ability to judge sociability from the voice as transmitted over a public address system. J. soc. Psychol., 1941, 13, 303-309. (a)

Fay, P. J. & Middleton, W. C. The ability to judge truth-telling, or lying, from the voice as transmitted over a public address system. J. gen. Psychol., 1941, 24, 211-215. (b)

Fay, P. J. & Middleton, W. C. Judgment of introversion from the transcribed voice. Quart. J. Speech, 1942, 28, 226-228.

Fay, P. J. & Middleton, W. C. Judgment of leadership from transmitted voice. J. soc. Psychol., 1943, 17, 99-102.

Fletcher, H. Speech and hearing in communication. New York: D. Van Nostrand, 1953.

REFERENCES (Continued)

French, N. R. & Steinberg, J. C. Factors governing the intelligibility of speech sounds. J. acoust. Soc. Amer., 1947, 19, 90-119.

Glauber, P. Speech characteristics of psychoneurotic patients. J. speech hear. Dist., 1944, 9, 18-30.

Goldfarb, W., Braunstein, Patricia & Lorge, I. A study of speech patterns in a group of schizophrenic children. Amer. J. Orthopsychiat., 1956, 26, 544-555.

Goldman-Eisler, Frieda. Speech-breathing activity—a measure of tension and affect during interviews. Brit. J. Psychol., 1955, 46, 53-63.

Goldman-Eisler, Frieda. Speech-breathing activity and content in psychiatric interviews. Brit. J. med. Psychol., 1956, 29, 35-48. (a)

Goldman-Eisler, Frieda. The determinants of the rate of speech output and their mutual relations. J. psychosom. Res., 1956, 1, 137-143. (b)

Graves, R. Watch the north wind rise. New York: Creative Age Press, 1949.

Greenfield, N. An experimental approach to vocal-verbal conflict. Percept. mot. Skills, 1958, 8, 251-262.

Hanawalt, N. Test review No. 28. In Buros, O. (ed). The fourth mental measurements yearbook. New Jersey: Gryphon, 1953.

REFERENCES (Continued)

Hargreaves, W. A. & Starkweather, J. A. Vocal behavior: an illustrative case study. Paper read at Western Psychological Association, Seattle, June, 1961. (a)

Hargreaves, W. A. & Starkweather, J. A. Drug effects on vocal behavior. Amer. Psychol., 1961, 16, 426. (b)

Harris, R. E. Dominance, assertiveness, and hostility in persons with high and low blood pressures: some similarities and some differences. Unpublished manuscript. Berkeley: Univ. of Calif., Institute of Personality Assessment and Research, 1953.

Herzog, H. Stimme und Persönlichkeit. Zsch. f. Psychol., 1933, 130, 300-379.

Isakower, O. On the exceptional position of the auditory sphere. Intern. J. Psychoanalysis, 1939, 20.

Jones, H. E. The adolescent growth study: VI. The analysis of voice records. J. consult. Psychol., 1942, 6, 255-256.

Kauffman, P. E. An investigation of some psychological stimulus properties of speech behavior. Unpublished Ph.D. dissertation, Univ. of Chicago, 1954.

Klopfer, B., Ainsworth, Mary, Klopfer, W. G. & Holt, R. R. Developments in the Rorschach technique. New York: World Book, 1954.

REFERENCES (Continued)

Knower, F. H. Analysis of some experimental variations of simulated vocal expressions. J. soc. Psychol., 1941, 14, 369-372.

Koch, Helen L. Sibling influence on children's speech. J. speech Dis., 1956, 21, 322-328.

Krause, M. S. Anxiety in verbal behavior: an intercorrelational study. J. consult. Psychol., 1961, 25, 272.

Krause, M. S. & Pilisuk, M. Anxiety in verbal behavior: a validation study. J. consult. Psychol., 1961, 25, 414-419.

Kretschmer, E. Physique and character. New York: Harcourt, 1925. (Translated by W.J.H. Spratt from Körperbau und Charakter. Berlin: Springer, 1921.)

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In Rubinstein, E. A. & Parloff, M. B. (eds). Research in psychotherapy. Washington, D.C.: APA, 1959, Pp. 160-208.

Licklider, J.C.R. & Miller, G. A. The perception of speech. In Stevens, S. S. (ed). Handbook of experimental psychology. New York: Wiley, 1951, Pp. 1040-1074.

REFERENCES (Continued)

Lynch, G. A phonographic study of trained and untrained voices reading factual and dramatic material. Archives of Speech, 1934, 1, 9-25.

Mahl, G. F. Disturbances and silences in the patient's speech in psychotherapy. J. abnorm. soc. Psychol., 1956, 53, 1-15.

Mahl, G. F. Measuring the patient's anxiety during interview from "expressive" aspects of his speech. Trans. N. Y. Acad. Sci., 1959, 21, 249-257.

Mallory, E. & Miller, V. A possible basis for the association of voice characteristics and personality traits. Speech Monogr., 1958, 255-260.

Malmo, R. B., Boag, T. J. & Smith, A. A. Physiological study of personal interaction. Psychosom. Med., 1957, 19, 105-119.

Maslow, A. H. Dominance-feeling, behavior, and status. Psychol. Rev., 1937, 44, 404-429.

McKelvey, D. P. Voice and personality. Western Speech, 1953, 17, 91-94.

McPherson, M. W. Speech behavior and egocentricity. J. clin. Psychol., 1956, 12, 229-235.

McQuown, N. A. Linguistic transcription and specification of psychiatric interview materials. Psychiatry, 1957, 20, 79-86.

Michael, W. & Crawford, C. C. An experiment in judging intelligence by the voice. J. educ. Psychol., 1927, 18, 107-114.

REFERENCES (Continued)

Monrad-Krohn, G. H. Dysprosody or altered "melody of language." *Brain*, 1947, 70, 405-415.

Monrad-Krohn, G. H. The third element of speech: prosody in the neuropsychiatric clinic. *J. ment. Sci.*, 1957, 103, 326-331.

Moore, W. E. Personality traits and voice quality deficiencies. *J. speech Dis.*, 1939, 4, 33-36.

Moses, P. Social adjustment and voice. *Quart. J. Speech*, 1941, 27, 532-536.

Moses, P. The study of personality from records of the voice. *J. consult. Psychol.*, 1942, 6, 257-261.

Moses, P. *The voice of neurosis*. New York: Grune & Stratton, 1954.

Moskowitz, Estelle W. Voice quality in the schizophrenic reaction type. *Speech Monogr.*, 1952, 19, 118-119.

Ochai, Y. & Fukumura, T. On the fundamental qualities of speech in communication. *J. acoust. Soc. Amer.*, 1957, 29, 392-393.

Ortleb, R. An objective study of emphasis in oral reading of emotional and unemotional material. *Speech Monogr.*, 1937, 4, 56-69.

Ostwald, P. A method for the objective denotation of the sound of the human voice. *J. psychosom. Res.*, 1960, 4, 301-305.

REFERENCES (Continued)

Paul, J. E. An investigation of parent child relationships in speech: intensity and duration. Unpublished Ph.D. dissertation, Purdue Univ., 1951.

Pear, T. H. Voice and personality. London: Chapman and Hall, 1931.

Pear, T. H. Speech as an expression of personality. Brit. J. educ. Psychol., 1932, 2, 139-152.

Pfaff, P. L. An experimental study of communication of feeling without contextual material. Speech Monogr., 1954, 21, 155-156.

Pittinger, R. E. & Smith, H. L. A basis for some contributions of linguistics to psychiatry. Psychiatry, 1957, 20, 61-78.

Pollack, I., Rubenstein, H., & Horowitz, A. Recognition of "verbal expression" over communications systems. Paper delivered at the 59th meeting of the Acoustical Society of America, 1960. (a)

Pollack, I., Rubenstein, H., & Horowitz, A. Communication of verbal modes of expression. Language and Speech, 1960, 3, 121-130. (b)

Potter, R. K., Kopp, G. A. & Green, H. C. Visible speech. New York: Van Nostrand, 1947.

Ramm, Katharine. Is monotonism an indication of maladustment? Smith College Studies in Social Work, 1946, 17, 264-284.

REFERENCES (Continued)

Rubinstein, E. A. & Parloff, M. B. (eds). Research in psychotherapy. Washington, D.C.: APA, 1959.

Sanford, F. H. Speech and personality. Psychol. Bull., 1942, 39, 811-845.

Sapir, E. Speech as a personality trait. Amer. J. Sociol., 1927, 32, 892-905.

Schilling, R. Über die Stimme erbgleicher Zwillinge. Folia Phonist., Basel, 1950, 2, 98-119; 205-238.

Scott, W.C.M. Noise, speech and technique. Int. J. Psychoanal., 1958, 39, 108-111.

Shakow, D. Discussion of papers by Leary & Gill, and Rogers. In Rubinstein, E. A. & Parloff, M. B. (eds). Research in psychotherapy. Washington, D.C.: APA, 1959. Pp. 108-115.

Skinner, E. R. A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness. Speech Monogr., 1935, 2, 81-137.

Soskin, W. F. Some aspects of communication and interpretation in psychotherapy. Paper read at APA, Cleveland, September, 1953.

Soskin, W. & Kauffman, P. Judgment of emotion in word-free voice samples. J. Communication, 1961, 11, 73-80.

Spilka, B. Relationships between certain aspects of personality and some vocal effects of delayed speech feedback. *J. speech hearing Disorders*, 1954, 19, 491-503.

Stagner, R. Judgments of voice and personality. *J. educ. Psychol.*, 1936, 27, 272-277.

Stagner, R. The Wisconsin scale of personality traits. *J. abnorm. soc. Psychol.*, 1937, 31, 463-471.

Starkweather, J. A. The communication value of content-free speech. Paper read at Midwest. Psychol. Assoc., Chicago, April, 1955. (a)

Starkweather, J. A. Judgments of content-free speech as related to some aspects of personality. Unpublished Ph.D. dissertation, Northwestern Univ., 1955. (b)

Starkweather, J. A. The communication-value of content-free speech. *Amer. J. Psychol.*, 1956, 69, 121-123. (a)

Starkweather, J. A. Content-free speech as a source of information about the speaker. *J. abnorm. soc. Psychol.*, 1956, 52, 394-402. (b)

Starkweather, J. A. Vocal communication of personality and human feelings. *J. Communication*, 1961, 11, 63-72.

REFERENCES (Continued)

Sullivan, H. S. The psychiatric interview. New York: Norton, 1954.

Taylor, H. C. Social agreements on personality traits as judged from speech. J. soc. Psychol., 1934, 5, 244-248.

Taylor, Janet. A personality scale of manifest anxiety. J. abnorm. soc. Psychol., 1953, 48, 285-290.

Terman, L. M. Terman group test of mental ability. Yonkers, N. Y.: World Book, 1920.

Tew, R. E. The relationship between the vocal characteristics of men and their ratings of the vocal characteristics of other men. Speech Monogr., 1956, 23, 26-30.

Thompson, C. W. & Bradway, Katherine. The teaching of psychotherapy through content-free interviews. J. consult. Psychol., 1950, 14, 321-323.

Thurstone, L. L. & Thurstone; Thelma G. Personality schedule. Chicago: Univ. of Chicago, 1929.

Tyler, Leona. Test review No. 77. In Buros, O. (ed). The fourth mental measurements yearbook. New Jersey: Gryphon, 1953.

Wagoner, L. C. Speech as an indication of temperamental traits. Quart. J. speech Educ., 1925, 11, 237-242.

REFERENCES (Concluded)

Wagoner, L. C. & Downey, J. E. Speech and will temperament. *J. appl. Psychol.*, 1922, 6, 291-297.

Wolff, W. *The expression of personality*. New York: Harper, 1943.

Zucker, L. Psychological aspects of speech-melody. *J. soc. Psychol.*, 1946, 23, 73-128.

Footnote

1. This paper was prepared in part during the tenure of a Predoctoral Fellowship from the National Institute of Mental Health, United States Public Health Service, and in part under Contract No. SAE-9265 with the Language Development Section, United States Office of Education.

The author wishes to express his appreciation to Harlan Lane and Alvin Burstein for reading the manuscript and offering many pertinent criticisms and suggestions.

PERSONALITY STEREOTYPES IN VOICE:
A RECONSIDERATION OF THE DATA

Ernest Kramer
The University of Michigan

Personality Stereotypes in Voice:
A Reconsideration of the Data[1]

Ernest Kramer
University of Michigan

The words "person" and "personality" derive from the Latin, *personare*, "to sound through" (*Oxford English Dictionary*, 1933, Vol. VII, p. 724). Apparently, the word referred to the mouth opening in the mask of an actor. Eventually the term shifted to mean the actor, himself, and then to mean any particular individual; but the etymological origin of "personality" is in the voice of the speaker (Moses, 1954).

A body of psychological research exists which has attempted to investigate experimentally what indications of personality may be found in the sound of the voice (Sanford, 1942; Licklider & Miller, 1951; Starkweather, 1961; Kramer, 1962). Typically, judges have listened to a group of unseen speakers and attempted to match the voices with a list of personality traits. When Sanford reviewed such studies in 1942, he felt that only slight relationships between voice and personality had been experimentally established. Common experience, however, seemed to verify the connection so strongly that he concluded, if the experimental approach "reveals no relationship, we would be forced to conclude that it may be the fault of the approach" (Stanford, 1942, p. 838). Starkweather, (1961), noting in a recent review the failure of many studies to demonstrate a relationship between judgments from voice and other criteria of personality, was left "pessimistic concerning the utility of assessing such traits from nonverbal stimuli" (p. 65). He made special reference to the frequent finding that the listener-judges agree better with one another than they do with external criteria. This finding has generally been ascribed to the existence of stereotyped voices: voices which convey a stereotype of some personality trait without having any actual validity

(Sanford, 1942; Starkweather, 1961). The present paper suggests that this inter-judge agreement is not without validity, and that the role of seeking correlations with external criteria has not been fully understood in such studies.

Of the many studies in this area, only two (Taylor, 1934; Fay & Middleton, 1940) found a tendency for judges to agree more frequently when they were in actual disagreement with other criteria, and in neither case was this tendency statistically significant. More typically, studies have found either inter-judge agreement and only "chance" correlations with the external criteria (Pear, 1931; Stagner, 1936; Fay & Middleton, 1941, 1942, 1943), or inter-judge agreement plus significant correlations with other criteria (Pear, 1931; Allport & Cantril, 1934; Cantril & Allport, 1935; Eisenberg & Zalowitz, 1938; Fay & Middleton, 1940). Even this last group of studies has concluded that vocal stereotypes exist and are invalid, because the correlations with external criteria have not been as great as those between judges. Stereotypes seem to be regarded as necessarily invalid. One study (Allport & Cantril, 1934) even used them to explain away some unexpectedly accurate listener judgments. What is it, then, that judges are reliably judging? Only the early study by Pear (1934) gave any evidence concerning the origin of these vocal stereotypes. His data suggest that some of them are due to conventionalized theatrical portrayals; but the source of others is wholly untraced, and even the origin of the theatrical conventions is unknown.

The personality traits being judged in such studies, those traits for which some voices provide presumably erroneous stereotypes, are not defined by a set of laboratory operations. They come from common experience or expert judges' reactions to persons, as do most of our personality trait labels. Only part of any such personality construct is operationally defined by a test designed to measure it; part of the trait remains unmeasured (Cronbach

& Meehl, 1955). The validity criteria, such as the Bernreuter Personality Inventory (Bernreuter, 1931), which have been most frequently used in voice and personality studies, are often highly imperfect measures of those traits that they are used to validate (McKelvey, 1953; Tyler, 1953). A trait such as "introversion" might be as validly measured by judgments from voice as by a test scale, yet each method might cover different portions of the total variance due to the trait and thus show little correlation with each other. Campbell (1960) has given a description of trait validity which fits this situation, if the phrase "the judgments from voice" is substituted for the word, "test":

> "...no a priori defining criterion is available as a perfect measure or defining operation against which to check the fallible test... (The) independent measure has no status as the criterion for the trait, nor is it given any higher status for validity than is the test. Both are regarded as fallible measures, often with known imperfections, such as halo effects for the ratings and response sets for the test. Validation, when it occurs, is symmetrical and equalitarian." (pp. 547-548)

Seen within this framework, the listener judgments are as valid a measure of a trait as are the test scores which have been used for the external criteria. Any positive correlation between them increases the presumptive validity of both.

Validity is best established by agreement between different and independent measurement procedures (Campbell, 1960; Campbell & Fiske, 1959). A single judge represents a single measurement procedure. If he repeatedly judges a personality trait as being present in a certain voice, he is--ignoring variations in conditions over time--merely establishing reliability and not validity. With several judges, however, each represents a somewhat different measurement procedure. The greater the number and heterogeneity of judges, the more agreement among them may be taken to represent validity, as well as reliability. Four sources of individual differences among listeners are noted further on in this paper; they are possible sources of meaningful heterogeneity. In general, studies such as that by Eisenberg and Zalowitz (1938), where the listeners were forty-three students in a psychology class, do not add as much towards establishing validity as do the judgments which Pear (1931) collected from over 4,000 radio listeners. In either case, presumptive validity is, of course, increased by positive correlations with other criteria.

Once it is seen that the presence of so-called "vocal stereotypes" is not really so empty a finding after all, several problems in the typical experimental approach to voice and personality do still remain. Although they are not the chief focus of this paper, two problems which have received virtually no mention in the literature may be briefly noted here.

First, most of the voice samples used have been monologues. The speakers have recited or read alone some standard passage. Many of the personality traits which listeners have tried to judge are ones usually associated with interactions between persons; dominance and submission, for example (Eisenberg & Zalowitz, 1938). The vocal cues for such traits seem more likely to appear in dialogue, such as might be gathered through role-playing scenes, than in monologue recitations and readings. This paper has concentrated on the relationship between voice and relatively stable personal characteristics, but consideration of voices interacting might be particularly useful in studies of how changing emotional states are indicated in the voice.

The second neglected area is also important for studies of changing emotional states. Various studies on voice have dealt with differences among speakers, but individual differences among listeners have been ignored. These differences among listener-judges may, as noted above, be capitalized upon to increase the validating power of inter-judge agreement. Four sources of such difference are (Kramer, 1962):

"(1) Personality variables. In addition to the effect these may have on the ability to make general personality judgments from voice, the literature on perceptual defense and need-motivated perception (Atkinson, 1958) suggests that the listener's motivational-need structure may strongly influence his judgments of particular characteristics in others.

"(2) Developmental variables. Theorizing has suggested that children may be relatively more sensitive than adults to the nonverbal aspects of speech (Soskin, 1953; Soskin & Kaufman, 1961), but no empirical evidence has been reported.

"(3) Psychophysical variables. How do individual differences in acuity to the various dimensions of sound affect listeners' perceptions of a speaker's personality and emotions?

"(4) Cultural-linguistic variables. In what way do the nonverbal cues (for personality and emotion) in speech vary from one language group to another?" (p. 44). Consideration of such differences should help to clarify the validating role of agreement among judges, as well as add to the general design of studies on personality and voice.

References

Allport, G. W., & Cantril, H., Judging personality from the voice. J. Soc. Psychol., 1934, 5, 37-55.

Atkinson, J. W., (Ed.), Motives in fantasy, action, and society. Princeton, N. J.: Van Nostrand, 1958.

Bernreuter, R., Personality inventory. Stanford, Calif.: Stanford Univ., 1931.

Campbell, D. T., & Fiske, D. W., Convergent and discriminant validation by the multitrait-multimethod matrix. Psychol. Bull., 1959, 56, 81-105.

Campbell, D. T., Recommendations for APA test standards regarding construct, trait, or discriminant validity. Amer. Psychol., 1960, 15, 546-553.

Cantril, H. & Allport, G., The psychology of radio. New York: Harper, 1935.

Cronbach, L. J., & Meehl, P. E., Construct validity in psychological tests. Psychol. Bull., 1955, 52, 281-302.

Eisenberg, P., & Zalowitz, E., Judgments of dominance-feeling from phonograph records of voice. J. appl. Psychol., 1938, 22, 620-631.

Fay, P. J., & Middleton, W. C., Judgment of intelligence from the voice as transmitted over a public address system. Sociometry, 1940, 3, 186-191.

Fay, P. J., & Middleton, W. C., The ability to judge sociability from the voice as transmitted over a public address system. J. soc. Psychol., 1941, 13, 303-309.

Fay, P. J., & Middleton, W. C., Judgment of introversion from the transcribed voice. Quart. J. Speech, 1942, 28, 226-228.

Fay, P. J., & Middleton, W. C., Judgment of leadership from transmitted voice. J. soc. Psychol., 1943, 17, 99-102.

Kramer, E., The judgment of personal characteristics and emotions from nonverbal properties of speech. Ann Arbor, Univ. of Mich. Office of Research Administration, report 04411-2-P, 1962.

Licklider, J. C. R., & Miller, G. A., The perception of speech. In Stevens, S. S., (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951, 1040-1074.

McKelvey, D. P., Voice and personality. *Western Speech*, 1953, 17, 91-94.

Moses, P. J., *The voice of neurosis*. New York: Grune & Stratton, 1954.

Oxford English Dictionary. Oxford: Clarendon Press, 1933.

Pear, T. H., *Voice and personality*. London: Chapman and Hall, 1931.

Sanford, F. H., Speech and personality. *Psychol. Bull.*, 1942, 39, 811-845.

Soskin, W. F., Some aspects of communication and interpretation in psychotherapy. Paper read at APA, Cleveland, September, 1953.

Soskin, W., & Kauffman, P., Judgment of emotion in word-free voice samples. *J. Communication*, 1961, 11, 73-80.

Stagner, R., Judgments of voice and personality. *J. educ. Psychol.*, 1936, 27, 272-277.

Starkweather, J., Vocal communication of personality and human feelings. *J. Communication*, 1961, 11, 63-72.

Taylor, H. C., Social agreements on personality traits as judged from speech. *J. soc. Psychol.*, 1934, 5, 244-248.

Tyler, Leona, Test review no. 77. In Buros, O. (Ed.), *The fourth mental measurements yearbook*. New Jersey: Gryphon, 1953.

Footnote

1. This paper was prepared during the tenure of a Predoctoral Fellowship from the National Institute of Mental Health, United States Public Health Service.

RESEARCH IN PROGRESS

REINFORCEMENT OF COVERT VOCAL RESPONDING

ON THE PERCEPTION OF AFFECT IN SPEECH

CONTROL OF AUDITORY ATTENTION

SOME PROPERTIES OF COVERT CHAINING IN LEARNING TO IDENTIFY STIMULI

THE PERCEPTION OF AFFECT IN A FOREIGN LANGUAGE AS A FUNCTION OF SECOND-LANGUAGE TRAINING

Research in Progress

1. Reinforcement of covert vocal responding.

In a previous study from this laboratory, Cross and Lane (1961) conditioned two concurrent vocal responses to two intensities of a pure tone and then tested for auditory generalization. Reciprocal generalization gradients, obtained for the two responses, partitioned the stimulus continuum into two classes. Within each class, the stimuli were equally effective in controlling response probability. Response latency was found to be inversely related to the ratio of the response probabilities at each stimulus intensity. The present study employs essentially the same experimental paradigm; however, an additional phase is interpolated between discrimination training and generalization testing. In this phase, the experimental group is exposed to stimuli adjoining the discriminative stimuli of phase 1 and, after a pause, the reinforcing stimulus is also presented. The subject is instructed to say nothing aloud, so that reinforcement is not response contingent. However, covert responding may be under discriminative control and may be reinforced. A control group receives the same treatment but is instructed to respond overtly in phase 2. A comparison of the generalization gradients for the experimental and control groups should reveal the efficacy, if any, of "differential reinforcement" that is not contingent on overt behavior.

2. On the perception of affect in speech.

Passages designed to portray five different emotions have been recorded by actors. Each passage contains a 27-word section common to all passages. These common sections will form the basic stimulus materials for the four parts of the current research:

a. Listeners are asked to identify the affect of these passages while listening to normal recordings and to recordings on which high frequencies have been attenuated.. The differences in inter-listener reliability for the two conditions are taken as a measure of the amount of affective information which is lost by low-pass filtering--a process used in several studies of nonverbal affective communication in speech.

b. Listener ratings of the affect from normal recordings are compared with ratings from recordings in which the speech intensity is held constant. The difference is taken as a measure of the contribution of varying intensities to affective communication in speech.

c. Each listener's identitication of the affective content of the common passage presented in typescript (prior to listening) is compared with his ratings of the various recorded samples. As the affective content of the words alone is ambiguous, it is assumed that their ratings reflect individual differences in attitude or emotion among the listeners; it is expected that these differences will be reflected in subject ratings of the recorded samples.

d. The stimulus passages are then recorded by those listeners whose judgments were most and least typical, and the common excerpts from these presented to a new group of listeners. It is hypothesized that those who are most typical in their judgments of affect in the speech of others are the more successful at producing recognizable affect in their own speech.

3. Control of auditory attention.

The effect of schedules of reinforcement on selective auditory attention is examined by controlling an overt correlate of attention. In this respect the research parallels that of J. G. Holland on the control of visual observing behavior. However, the experimental paradigm is extended to the concurrent responding situation, as it is desired to examine the effects of competing contingencies of reinforcement on incompatible responses.

The subject is provided with a dichotic binaural headset and two manipulanda, one controlling each earphone. Depression of either button connects the corresponding earphone to an audio source for two seconds. In an initial study, random numbers are recorded on each track of a two-channel tape recording. The reinforcing stimulus (a selected number that S is instructed to report) is interpolated in each series at four-minute intervals. Two concurrent four-minute fixed interval schedules, with limited hold, are therefore in effect. The dependent variables are the time distributions of responding on each button and switching between buttons. A major experimental parameter is the "phase angle" of the two FI schedules, that is, the relative spacing of the reinforcers.

In order to obtain data on S's covert responding while attending to one source, and to preclude counting as a means of "timing" the fixed interval, a probe technique is employed. At a selected moment during certain of the fixed intervals, a light flash is presented and S then reports as many of the preceding train of numbers as he can. Per cent recall while monitoring one source may be a function of the location of the probe during the fixed interval.

4. Some properties of covert chaining in learning to identify stimuli.

The behavior of "identifying" each of a set of stimuli involves (1) stimulus discrimination, (2) response differentiation, (3) the coordination of the first two repertories. Keller and Schoenfeld (1950) have described the development of the coordination of the two repertories in code-receiving, which may serve as a model for many "language skills": (the description assumes that the student has previously learned to discriminate the temporal properties of auditory signals and that the necessary responses were differentiated when he learned to write his language)

"In the first stage of code receiving, the signal occasions various

> responses, sometimes overt, which serve in turn as stimuli for the copying response that ends the sequence. Later, these intervening responses become covert, although still present as members of the chain. Finally, they are eliminated entirely, and the observed decrease in latency is thereby made possible." (p. 218)

The present study examines some properties of the vanishing intervening behavior in learning to identify auditory stimuli. The stimulus set consists of one-second trains of pulses, varying in repetition rate from 6 to 18 pps. The response set consists of estimates of the number of pulses in a train. Correct identifications are reinforced with points. Correct and incorrect responses and their latencies are recorded. Early in training, stimuli with lower pps values evoke counting prior to identification; of course, stimuli with high pps values do not. Thus, the conditioning of stimulus identification with and without intervening chaining may be examined under conditions that are comparable in other respects. Response latency is related to the time course of acquisition under both conditions, and intervening stimulus values are employed to measure stimulus generalization. Various training and testing procedures are under study.

5. The perception of affect in a foreign language as a function of second-language sophistication.

In a review of the experimental literature on the judgment of personal characteristics and emotions from the nonverbal properties of speech, Kramer (1962) comments: "As a whole, there is support for the popular notion that one can tell something about a person and his feelings from the sound of his voice and the manner of his speaking." It is interesting to inquire whether there are cross-linguistic invariances in the encoding of affect in speech and whether formal instruction in a second language enhances perception of the nonverbal cues to affect in that language. In an initial study, passages that appear neutral in content have been excerpted from Japanese psychotherapy recordings. This neutrality is validated by affect ratings of transcriptions of the passages. A subset of the recordings is then presented to native Japanese for affect rating. Prior research indicates that appreciable inter-judge agreement may be anticipated (Kramer, 1961). These native stereotypes serve as criteria for estimating the amount of adventitious learning to discriminate affect cues that is associated with various amounts of instruction in Japanese. Invariance in the affect code, in this case across language families, is estimated from a comparison of affect ratings by Americans with no training in Japanese and the criterion stereotypes.

References

Cross, D. V. and Lane, H. L. On the discriminative control of concurrent responses: the relations among response frequency, latency, and topography in auditory generalization. In: *Experimental analysis of the control of speech production and perception*. Ann Arbor: University of Michigan research report 04411-1-P, 1961.

Holland, J. G. Human vigilance. *Science*, 1958, 128, 61-67.

Keller, F. S. and Schoenfeld, W. N. *Principles of psychology*. New York: Appleton-Century, 1950.

Kramer, E. The judgment of personal characteristics and emotions from the nonverbal properties of speech. In: *Experimental analysis of the control of speech production and perception*: II. Ann Arbor: University of Michigan research report 04411-2-P, 1962.

APPENDIX

STAFF

SPECIAL ACTIVITIES

DISTRIBUTION

Staff

During the past five months, the following have been members of the research and administrative staff of the project.

H. L. Lane, Ph.D., Director
D. V. Cross, M.A., Research Assistant
E. F. Kramer, M.A., Research Assistant
P. G. Shinkman, Ph.D., Research Assistant
R. L. Buiten, B.S.E., Research Engineer
D. R. Brinkman, Technical Assistant
B. A. Schneider, Technical Assistant
C. M. Hogue, Secretary

The invaluable assistance of the staff and facilities of the Communication Sciences Laboratory (G. E. Peterson, Ph.D., Director) is gratefully acknowledged.

Special Activities

Addresses presented or scheduled by project personnel:

1. Shaping the prosodic features of speech with an autoinstructional device. Amer. Assn. for the Adv. of Science, Denver, 1961. (H. L. Lane)
2. The effects of changing vowel parameters on perceived loudness and stress. Acoust. Soc. Amer., Cincinnati, 1961. (H. L. Lane)
3. Specifications for auditory discrimination learning in the language laboratory. Indiana-Purdue Lang. Lab. Conf., Bloomington, 1962. (H. L. Lane)
4. Programmed language learning. Northeast Conference on Language Learning, Boston, 1962. (H. L. Lane)
5. Methods for self-shaping echoic behavior. Eastern Psychol. Assn., Atlantic City, 1962. (H. L. Lane and B. A. Schneider)
6. Discriminative control of concurrent responding. Eastern Psychol. Assn., Atlantic City, 1962. (D. V. Cross and H. L. Lane)
7. Parameters of vowel perception. Acoust. Soc. Amer., New York, 1962. (H. L. Lane)
8. Influence of verbal conditioning on non-verbal behavior. Psychonomic Society, St. Louis, 1962. (H. L. Lane)
9. Personality stereotypes in voice. Michigan Psychol. Assn., Detroit, 1962. (E. M. Kramer)

Distribution List

Dr. A. B. Gaarder
Research and Studies Unit
Language Development Section
Division of Higher Education
Office of Education
Washington 25, D. C.

Dr. D. A. Cook
Basic Systems Incorporated
42 East 52nd Street
New York 22, New York

Dr. W. S-Y Wang
Denney Hall
Ohio State University
Columbus, Ohio

Dr. Dennis Fry
Department of Phonetics
University of London
London, England

Dr. A. Liberman
Haskins Laboratories
305 East 43rd Street
New York, New York

Dr. Gunnar Fant
Kungl. Tekniska Hogskolan
Institutionen for Telegrafi-Telefoni
Stockholm 70, Sweden

Dr. F. R. Morton
Institute for Behavioral Research
and Programmed Instruction
1315 Hill Street
Ann Arbor, Michigan

Dr. J. V. McConnell
Britannica Center
P. O. Box 11576, Station A
Palo Alto, California

Dr. A. Hayes
715 Gilbert
Tacoma Park 12, Maryland

Dr. R. W. Brown
Department of Psychology
Massachusetts Institute of Technology
Cambridge, Massachusetts

Dr. W. McGill
Department of Psychology
Schermertorn Hall
Columbia University
New York 27, New York

Dr. B. F. Skinner
Department of Psychology
Memorial Hall
Harvard University
Cambridge 38, Massachusetts

Dr. S. S. Stevens
Department of Psychology
Memorial Hall
Harvard University
Cambridge 38, Massachusetts

Mr. N. J. Peterson
Department of Psychology
Memorial Hall
Harvard University
Cambridge 38, Massachusetts

Dr. R. Glaser
Department of Psychology
University of Pettsburgh
Pittsburgh, Pennsylvania

Dr. W. N. Schoenfeld
Department of Psychology
Schermerhorn Hall
Columbia University
New York 27, New York

Dr. E. L. Walker
Department of Psychology
University of Michigan
Ann Arbor, Michigan

Dr. W. J. McKeachie
Department of Psychology
University of Michigan
Ann Arbor, Michigan

Mr. J. E. Lesch
Office Research Administration
University of Michigan
Ann Arobr, Michigan

Mr. Dale Brethower
Language Laboratory
University of Michigan
Ann Arbor, Michigan

Dr. Ilse Lehiste
Communication Sciences Laboratory
University of Michigan
Ann Arbor, Michigan

Mr. Arnold Roe
Teaching Systems Research Project
Department of Engineering
University of California
Los Angles 24, California

Mr. G. L. Geis
Psychology Department
Hamilton College
Clinton, New York

Dr. Aaron S. Carton, Head
Foreign Language Research Section
Experimental Teaching Center
New York University
School of Education
Washington Square, New York 3, N. Y.

Dr. L. S. Harms
Louisiana State University
Department of Speech
Baton Rouge 3, Louisiana

Dr. Evan R. Keislar
Department of Education
University of California
Los Angeles 24, California

Dr. John A. Barlow
Emory University
Department of Psychology
Atlanta 22, Georgia

Dr. John B. Carroll
Harvard University
Cambridge 38, Massachusetts

Dr. Howard L. Nostrand
Department of Romance Language
University of Washington
Seattle 5, Washington

Dr. Peter Ladefoged
Department of Phonetics
University of Edinburg
Edinburg, Scotland

Mr. Upshur
English Language Institute
University of Michigan
Ann Arbor, Michigan

Dr. J. A. Starkweather
Langley Porter Neuropsychiatric Institute
401 Parnassus Avenue
San Francisco 22, California

Mr. S. Margolis
Basic Systems, Inc.
42 East 52nd Street
New York 22, New York

Dr. R. S. Tikofsky
Communication Sciences Laboratory
University of Michigan
Ann Arbor, Michigan

Mr. Mark Rilling
3422 Tuland Drive
West Hyattsville, Maryland

Dr. Finley Carpenter
University High School
University of Michigan
Ann Arbor, Michigan

Professor Barrs
Electronics Research Laboratory
Northeastern University
Boston 15, Massachusetts

Dr. John W. Black
Department of Speech
Ohio State University
Columbus 10, Ohio

Mr. Brent Baxter
Director of Agencies Research
The Prudential Insurance Company of America
Newark, New Jersey

Mrs. Anita McMillan
Communications Research Center
Michigan State University
East Lansing, Michigan

Mrs. R. A. Glaser, Librarian
Gerber Baby Foods
445 State Street
Fremont, Michigan

Mrs. Sandra Gates
Research Department
Civic and Bovernmental Affairs
Ford Motor Company
Dearborn, Michigan

Mr. N. A. Fattu
Director, Institute of Educational Research
School of Education
Indiana University
Bloomington, Indiana

Dr. John Douglas
School of Business Administration
Wayne State University
Detroit 2, Michigan

Miss Ethel Halcomb
16174 Lawton Avenue
Detroit 21, Michigan

Mr. Bert Minwegen
Special Projects
Central Cooperatives, Inc.
Superior, Wisconsin

Mr. Leonard C. Gradeck
5806 Granklin Bldve.
Cleveland 2, Ohio

Mr. Francis C. Walker
Harless and Kirkpatrick Associates, Inc.
Professional Arts Building
420 West Lafayette Street
Tampa, Florida

Dr. Leslie G. Moeller
Director, School of Journalism
State University of Iowa
Iowa City, Iowa

Dr. Eleanor J. Gibson
Department of Psychology
Cornell University
Ithaca, New York

Dr. James N. Mosel
Department of Psychology
George Washington University
Washington 6, D. C.

Dr. Richard A. Littman
University of Oregon
Department of Psychology
Eugene, Oregon

Mr. Daniel J. Ryan
General Superintendent
City of Detroit
Department of Public Welfare
735 Randolph Street
Detroit 26, Michigan

Mr. Charles Colenaty
President, Counselors to Management, Inc.

3440 Wilshire Boulevard
Los Angeles 5, California

Dr. Lawrence Schlesinger
George Washington University
Washington 6, D. C.

Miss Helen G. Gommar, Head
Technical Information Section
Celanese Fibers Company
Box 1414
Charlotte 1, North Carolina

Mr. Sidney Coyne
3990 Saxon Avenue
Bronx 63, New York

Dr. R. M. Bourne
Business Organization and Management
University of Nebraska
College of Business Adminstration
Lincoln 8, Nebraska

Dean Mahlon H. Hellerich
Albright College
Reading, Pennsylvania

Mr. Calvin Otto
Prentice-Hall, Inc.
Englewood Cliffs, New Jersey

Mr. Lawrence E. Kanous
Employment and Personnel Research Dept.
The Detroit Edison Company
2000 Second Avenue
Detroit 26, Michigan

The Library
Reference Department
Humbolt State College
Arcata, California

Dr. Frank P. Piscor
Vice President of Academic Affairs
Syracuse University
Syracuse 10, New York

Dr. H. M. Wisely
Director, Personnel and Public Relations
Eli Lilly International Corporation
Indianapolis 6, Indiana

Miss Elizabeth M. Adles, Supervisor
Cook County Department of Public Aid
69 West Washington Street
Chicago 2, Illinois

Dr. Frank D. Horvay
Head, Department of Foreign Languages
Illinois State Normal University
Normal, Illinois

Mr. J. Hammock
Bell Telephone Laboratories, Inc.
Murray Hill, New Jersey

Mr. E. A. Roselene
Assistant General Manager
Planning and Development Department
Prudential Insurance Company of America
Newark, New Jersey

Mr. Eugene H. Rocklyn
Human Resources Research Office
George Washington University
Washington 7, D. C.

Mr. Hillman Yowell
Production Manager
California Date Growers Association
P. O. Box 577
Indio, California

Mr. D. G. Clough
Research Education
International Business Machines
Thoman S. Watson Research Center
P. O. Box 218
Yorktown Heights, New York

Lt. Paul G. Shinkman
U. S. Army Chemical Center
Edgewater, Maryland

Mrs. Mary L. Lindner
Ordnance Corps
Diamond Ordnance Fuze Laboratories
Washington 26, D. C.

Dr. Peter F. Ostwald
Langley Porter Neuropsychiatric Institute
401 Parnassus Avenue
San Francisco 22, California

Mr. Victor Sadler
Department of Phonetics
University of London
London England

Mr. J. E. Coulson
Systems Development Corp.
Santa Monica, California

Dr. Richard S. Crutchfield
Director, Research Study of Programmed
Instruction and Creative Thinking
University of Personality Assessment and Research
Berkeley 4, California

Mr. Charles D. Atkinson, III
Teaching Machine Project
Psychological Laboratories
Cambridge 38, Massachusetts

Mr. Johnnie E. Andre
2451 LeConte Avenue
Berkeley 9, California

Dr. Robert Shafer
College of Education
Wayne State University
Detroit 2, Michigan

Dr. Jean Malmstrom
Department of English
Western Michigan University
Kalamazoo, Michigan

Mr. George G. Slutsky
P. O. Box 294
Institute, West Virginia

Dr. Marjorie C. Johnston
5411 Trent Street
Chevy Chase 15, Maryland

Mr. Warren B. O'Neill
3833 Boudinct Avenue
Cincinnati 11, Ohio

Mr. G. Alonzo Stanford
3 Pooks Hill Road, Apt. 919
Bethesda 14, Maryland

Mr. J. Richard Reid
Department of Romance Languages
Clark University
Worcester 10, Massachusetts

Mr. Stanley M. Tsuzaki
1647-23 Beal Avenue
Ann Arbor, Michigan

Mr. Dwight L. Bolinger
University of Colorado
Boulder, Colorado

Dr. Edward Kupka
1501 New Hampshire NW
Washington 6, D. C.

Mr. William E. Norris
English Language Institute
University of Michigan
Ann Arbor, Michigan

Mr. Charles S. Weber
Civil Service Commission
320 South Walnut Street
Lansing, Michigan

Dr. Norman Markel
Dept. of Anthropology and
Linguistics
The University of Buffalo
Buffalo 14, New York

Dr. Wilson B. Bishai
School of Advanced International Studies
Johns Hopkins University
Washington 9, D. C.

Mr. T. A. Chapekis, Supervisor
Personnel & Organization Planning Dept.
Ford Motor Company
P. O. Box 2053
20000 Rotunda Drive
Dearborn, Michigan

Dr. Jerome D. Schein
Gallaudet College
Office of Psycho-Educational Research
Washington 2, D. C.

Mr. Edward C. Kilian
State Savings Bank Building
Frankfort, Michigan

Miss Joan Sayre
Department of Speech
Miami University
Oxford, Ohio

Professor H. H. Josselson
Department of Slavic and Eastern
Languages
Wayne State University
Detroit 2, Michigan

Dr. Donald Shearn
Department of Psychology
Colorado College
Colorado Springs, Colorado

Dr. Werner J. Koppitz
Thomas J. Watson Research Center
P. O. Box 218
Yorktown Heights, New York

Mr. Paul F. Ryan
University of Michigan
Administration Building
Ann Arbor, Michigan

Mr. David A. Reibel
Air Force Language Program
Indiana University
Bloomington, Indiana

Dr. H. S. Gershman
Department of Romance Languages
University of Missouri
Columbia, Missouri

Mr. William F. Marquardt
School of Education
New York University
New York, New York

Miss Dorothy M. Godlewski
19648 Charest
Detroit 34, Michigan

Mr. Francis N. Laird
Department 3163, Bldg. 2
Autonetics
9150 E. Imperial Hwy.
Downey, California

Mr. Julius Pochelon
Kenower, MacArthur & Company
Ford Building
Detroit 26, Michigan

Mr. John N. Seaman
1504 Michigan National Tower
Lansing 8, Michigan

Mr. Fred Brengelman
Department of English
University of Wisconsin
Madison 6, Wisconsin

Mr. Leon B. Feddersen
Epsilon Sigma Alpha
Loveland, Colorado

Dr. Frances Ingemann
University of Kansas
Language Sound Rooms, Blake Annex
Lawrence, Kansas

Dr. Seymour Chatman
Department of Speech
University of California
Berkeley 4, California

Dr. Paul Pimsleur
The Ohio State University
Listening Center
Denney Hall
164 West 17th Avenue
Columbus 10, Ohio

Dr. Frank A. Rice
Center for Applied Linguistics
1346 Connecticut Avenue,N. W.
Washington 6, D. C.

Mr. Daniel R. Stein
Department of Anthropology and Linguistics
The University of Buffalo
Buffalo 14, New York

Mr. David E. Carter
Psychology Department
Mason Hall
The University of Michigan
Ann Arbor, Michigan

Mr. James N. Davis
33 West Grand Avenue
Chicago 10, Illinois

Mr. Jack H. Feller, Jr.
The Permanente Medical Group
2350 Geary Blvd.
San Francisco, Calif.

Mr. R. Simon
California Research Corp.
P. O. Box 446
La Habra, Calif.

Miss Clara N. Bush, Director
English for Foreign Students
Department of Speech & Drama

Dr. Jerome D. Schein
Office of Psycho-Educational Research
Gallaudet College
Washington 2, D. C.

Jack H. Feller, Jr.
The Permanent Medical Group
2350 Geary Boulevard
San Francisco, California

Mr. Lloyd N. Morrisett
Carnegie Corporation of New York
589 Fifth Avenue
New York 17, New York

R. Simon
California Research Corporation
P. O. Box 446
La Habra, California

Clara N. Bush, Director
English for Foreign Students
Department of Speech and Drama
Stanford University
Stanford, California

Fred W. Herrmann
Contra Cost Junior College District
P. O. Box 431
Martinez, California